VOLUME 16, NUMBER 2 AUGUST 1960

WORD

JOURNAL OF THE LINGUISTIC CIRCLE OF NEW YORK

GENERAL LIBRARY
AUG 29 1960
UNIV. OF N. DAKOTA

EDITORS

ROBERT AUSTERLITZ
WILLIAM DIVER
ANDRÉ MARTINET
URIEL WEINREICH

CONTENTS

ARTICLES

The Short Vowel Systems of Northern Switzerland. A Study in Structural Dialectology *William G. Moulton* 155

The Consonantal Alternations in the Slavic Declensions *Edward Stankiewicz* 183

A Note on Slavic Verbs of the Type *zějǫ : zîjati* *William R. Schmalstieg* 204

Stylistic Context *Michael Riffaterre* 207

Estimating Vocabulary Size *Alvar Ellegård* 219

OBITUARY

Henri François Muller *Pauline Taylor* 245

REVIEW ARTICLE

Linguistic Structures *W. Haas* 251

REVIEWS

A. I. Smirnickij, Leksikologija anglijskogo jazyka *D. S. Worth* 277

Jacob M. Landau, A Word Count of Modern Arabic Prose *Majed F. Sa'id* 285

W. Kaper, Kindersprachforschung mit Hilfe des Kindes *Jean Berko* 287

Armas Salonen *und* Paavo Siro, Studien zur neusumerischen Syntax *Raymond-Riec Jestin* 290

L. Kaiser (ed.), Manual of Phonetics *Martin Kloster Jensen* 292

NOTICE 315

Published by

THE LINGUISTIC CIRCLE OF NEW YORK

LINGUISTIC CIRCLE OF NEW YORK, INC.

Cercle Linguistique de New York

PRESIDENT: **Elliott Van Kirk Dobbie,** Columbia University

VICE-PRESIDENTS: **Robert A. Fowkes,** New York University and Columbia University
André Martinet, Sorbonne and Ecole pratique des Hautes Etudes
Joshua Whatmough, Harvard University

SECRETARY and TREASURER: **John P. Hughes,** St. Peter's College and Columbia University

EXECUTIVE COMMITTEE: The preceding and

Marshall D. Berger, City College of New York
William Diver, Columbia University
Eugene Dorfman, University of Washington
Paul L. Garvin, Georgetown University
Judah A. Joffe, Ecole libre des Hautes Etudes
Alphonse G. Juilland, University of Pennsylvania and Columbia University
Fredrick H. Jungemann, Columbia University
James Macris, Rutgers University
Mel Most, New York
Leo Pap, State University Teachers' College, New Paltz, N.Y.
George C. Pappageotes, Columbia University
Jeanne V. Pleasants, Columbia University
Uriel Weinreich, Columbia University
Francis J. Whitfield, University of California

ADVISORY COMMITTEE FOR FOREIGN COUNTRIES:

Alfredo Barrera Vásquez, Mexico
Emile Benveniste, France
Giacomo Devoto, Italy
J. R. Firth, England
Henri Frei, Switzerland
J. Duchesne-Guillemin, Belgium
B. Havránek, Czechoslovakia
Louis Hjelmslev, Denmark
Bertil Malmberg, Sweden
J. Mattoso Camara, Brazil
†**Paul Menzerath,** Germany
Anton Reichling, The Netherlands
James St. Clair-Sobell, Canada
Alf Sommerfelt, Norway
Antonio Tovar, Spain

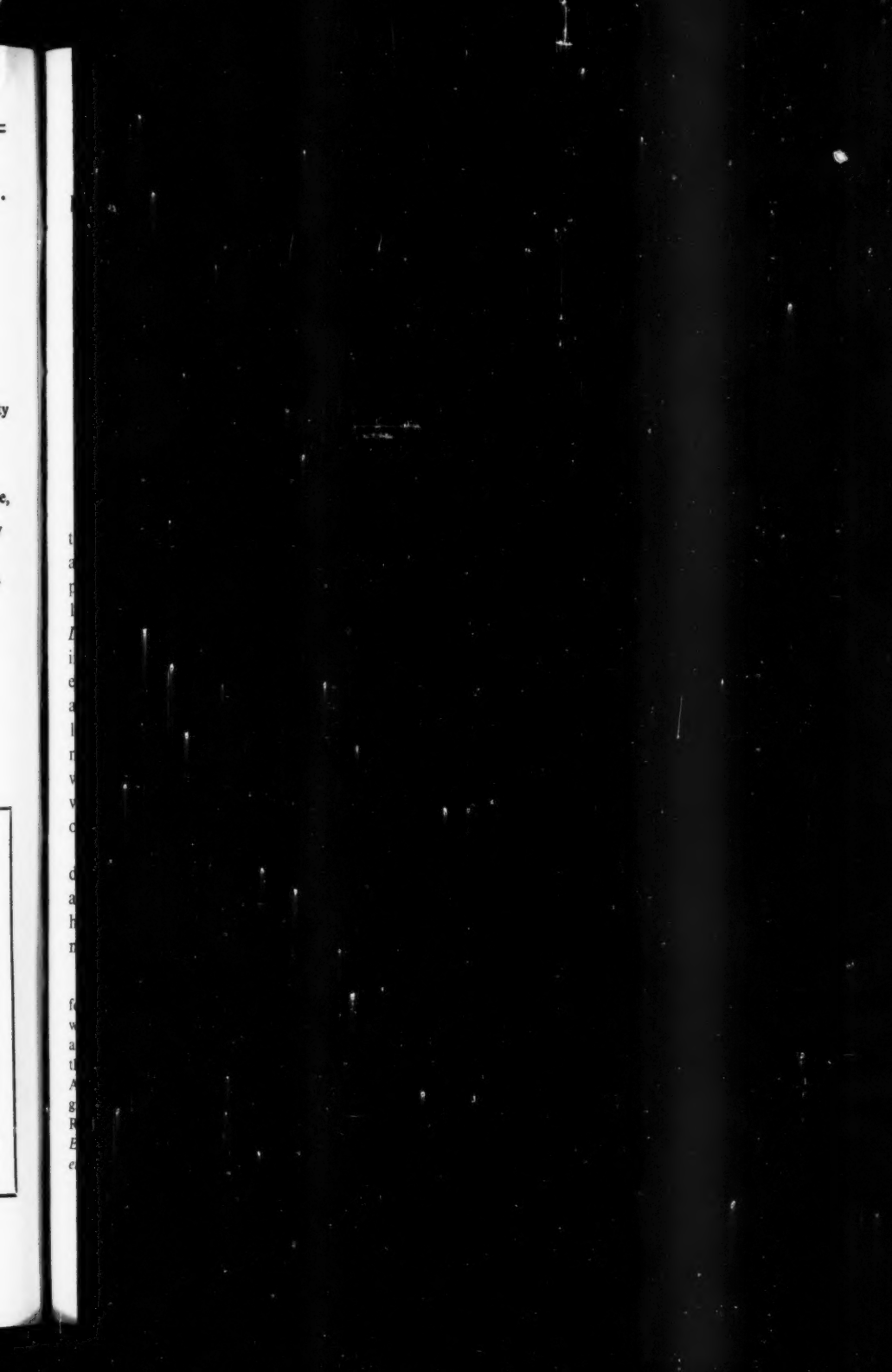
WORD, the journal of the Linguistic Circle of New York, is published three times a year in April, August, and December.

Subscription to WORD is $7.50 per year, this including membership in the Circle and postage anywhere in the world. For full-time students of linguistics, the subscription price is $5.00. Dues for membership in the Circle without subscription to the journal are $1.50 a year. Single numbers of the journal, $3.00 each.

Applications for membership, subscriptions to WORD, and orders for back numbers should be sent to the Secretary-Treasurer, **John P. Hughes,** St. Peter's College, 2641 Hudson Blvd., Jersey City 6, New Jersey.

Manuscripts of articles and monographs should be sent either to André Martinet (3, place de la Gare, Sceaux [Seine], France) or to William Diver or Uriel Weinreich (Box 14, Philosophy Hall, Columbia University, New York 27, N.Y., U.S.A.). **Books for review or listing** and **manuscripts of reviews** should be sent to William Diver at the above address. **Exchange journals** should be sent to André Martinet at the address given above.

The editors of the journal are jointly responsible for its contents. Authors of articles are allowed 100, and authors of reviews 25, reprints gratis; they may get additional copies at cost.



WILLIAM G. MOULTON

The Short Vowel Systems of Northern Switzerland

A Study in Structural Dialectology

The field records of a dialect atlas furnish valuable materials for testing the accepted theories of conventional phonemics, which deals with data assumed to be uniform, and for developing the new theories of dialect phonemics, which must deal with data clearly not uniform. During the year 1958–1959 the writer was privileged to work with the field records of the *Linguistic Atlas of German Switzerland* (*Sprachatlas der deutschen Schweiz*), in Zürich.[1] These are extensive records (200 to 300 handwritten pages each) of field interviews in 570 localities in German-speaking Switzerland and adjacent Italy, completed during a 20 year period ending in December, 1958. The speech of the informants was recorded impressionistically by means of a highly flexible set of phonetic symbols. No conscious attempt was made to normalize the transcription at any one locality; the field workers tried to note every shade of sound which they were able to record on paper.

The task of ordering such materials structurally is the familiar one of determining the phonemic relevance of phonetic contrasts. The methods are well known; they need not be described here. It is perhaps worth noting, however, that such a task is not lightly undertaken. My worksheets for the monophthongs of the Canton of Zürich (66 localities) contain well over

[1] I am deeply grateful to Prof. R. Hotzenköcherle, director and editor of the Atlas, for allowing me to use these materials with complete freedom. Through his kindness I was able to benefit from the years of labor which have already gone into the gathering and ordering of the materials, and to consult the many hand drawn maps which have thus far been prepared. My stay in Zürich was made possible through a grant from the American Council of Learned Societies, for which I would also like to express my gratitude. For a general description of the Atlas (publication has not yet begun), see R. Hotzenköcherle, "L'Atlas linguistique et ethnographique de la Suisse alémanique," *Essais de philologie moderne* (1951), pp. 115–32 (Paris, 1953; = *Bibl. de la Fac. de Phil. et Lettres de l'Univ. de Liège*, fascicule 129).

10,000 entries. Yet the gathering of data is of course the quickest part of the job. Most of the time is spent in trying to determine what data can and should be gathered, and then in subjecting the gathered data to analysis.

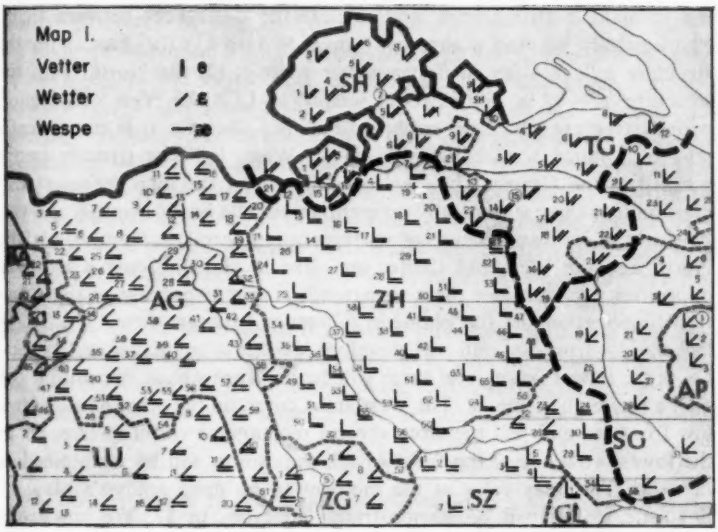
For the purposes of this paper there is chosen for study a section of northern Switzerland (about 78×108 kilometers or 48×67 miles in area) which is of special interest because of the different vowel structures which it shows. We limit our study to the short vowels; inclusion of the long vowels would expand it to monograph size. Since the 220 localities in this area are geographically contiguous, it is obvious that we shall wish to present our final data in map form. The question then arises: What structural features can be presented in this way, and what techniques should be used? At the one extreme we would like to present, for example, the whole vowel system: short vowels, long vowels, and diphthongs. A very practical difficulty makes this impossible. The area contains a great many such systems, and presenting them in map form would give so complicated a picture as to defeat the very purpose of cartographic representation. Further, one wants to know not merely that two vowel systems differ, but also how they differ; and not even the most skillful selection of symbols could show this.

At the other extreme one thinks of somehow plotting geographically certain data about a single phoneme. We hope to show later how this type of presentation can be made. Before it can be made, however, we shall need to know more about the full short vowel systems of the area. Only then will it be possible to say, for example, that a phoneme /e/ of one system is in some sense "the same" as a phoneme /e/ of another system. Obviously the two are different if they belong to different systems. Any "sameness" which they share must be more than the irrelevant fact that we symbolize them in the same way.

§1. FRONT UNROUNDED VOWELS. In our first three maps we have chosen to present more or less raw phonetic data, but to present them in such a way that a structural interpretation is immediately obvious. For Map 1 we selected three groups of five words each which we had reason to believe would allow us to set up three short vowel phonemes in part of the area, two in the rest.[2] The word groups chosen were: (1) *Vetter, Bett, Beck*

[2] Monographs on various parts of this area are available as follows: J. Hunziker, *Aargauer Wörterbuch in der Lautform der Leerauer Mundart*, Aarau, 1877 (Leerau is point AG 51 on our map); Heinrich Stickelberger, *Lautlehre der lebenden Mundart der Stadt Schaffhausen*, Leipzig diss., Aargau 1881 (Schaffhausen is point SH7); H. Blattner, *Über die Mundarten des Kantons Aargau: Vocalismus der Schinznacher Mundart*, Leipzig diss., Brugg, 1890 (Schinznach is near point AG 23); Ernst Hausknecht, *Die Vokale der*

(= *Bäcker*), *fest*, *gestern* (< Middle High German close *e*); (2) *Wetter*, *Speck*, *Fest*, *Nest*, *Rechen* (< MHG open *ë*); and (3) *Wespe*, *Pächter*, *Bäsi*, *(ein)tägig*, *färben* (< MHG *ä*). The phonetic values which the field workers gave for the stressed vowels of these words were entered on sheets of lined paper opposite the number of the appropriate locality. In most cases the



same value was given for all five words in each group. Where there were slight deviations, these were regarded as representing free allophonic variations of a single phoneme. In the rare cases where one word was markedly different from others in its group, it was separated from them, since its vowel was regarded as representing some other phoneme.[3] In this

Stammsilben in den Mundarten der Stadt St. Gallen und des Fürstenlandes, Zürich diss., Frauenfeld, [1912?] (St. Gall is five miles east of point SG 5); Ludwig Fischer, *Der Stammsilben-Vokalismus der Mundart des Luzerner Gäus*, Zürich diss., Frauenfeld, 1927 (based on the dialect of Triengen, point LU 5); Albert Weber, *Zürichdeutsche Grammatik*, Zürich, 1948; and the following volumes of the series *Beiträge zur Schweizerdeutschen Grammatik* (= *BSG*), ed. by Albert Bachmann, Frauenfeld, 1910ff.: *BSG* 1, Jakob Vetsch, Appenzell (1910); *BSG* 9, Wilhelm Wiget, Toggenburg (in St. Gall) (1916); *BSG* 15, Albert Weber, Zürich Highlands (1923); and *BSG* 20, Georg Wanner, canton of Schaffhausen (1941).

[3] For example, at AG 21 the word *Wespe* differs from the others in its group in showing close [e] rather than open [ɛ]. This [e] is assigned to the /e/ established by the words of group (1) rather than to the /ɛ/ established by groups (2) and (3).

way it was possible to arrive at an approximate phonetic value for the stressed vowel phoneme of each group at each locality. These values were then entered on the base map, with a vertical line representing higher mid [e], a diagonal line mean or lower mid [ɛ], and a horizontal line low front [æ]. The result is Map 1.[4]

Map 1 can be read either phonemically or phonetically. It will be interesting to make both readings, and to note the differences between them. Phonemically we find a three-way division into (1) the East, with the structure /e/—/ɛ/—/æ/ in /fettər wɛttər wæšpi/; (2) the North, with the structure /e/—/ɛ/ in /fettər wɛttər wɛšpi/; and (3) the West and Center, with a structure which we might note as /E/—/æ/, i.e. it is phonetically [ɛ]—[æ] ([fɛttər wættər wæšpi]) in the West, [e]—[æ] ([fettər wættər wæšpi]) in the Center. The isoglosses drawn on the map represent this division into three structurally different areas. The East obviously shows a structure of its own because of its three-way contrast /e/—/ɛ/—/æ/. The North and the West and Center are alike in showing only a two-way opposition; we consider them structurally different because of the structural implications of the phonetic difference between them. The North lacks a low front [æ] and will therefore show a triangular vowel system, i.e. at the lowest vowel level there will be no front ≠ back opposition but only a single phoneme /a/. The West and Center, on the other hand, have a low front [æ] and will therefore show a rectangular vowel system, i.e. at the lowest vowel level the front ≠ back opposition will be maintained as /æ/ ≠ /a/. This may seem at the moment to be mere analyst's play; its linguistic reality will be demonstrated later on, in §7. We tentatively consider the West with [ɛ]—[æ] and the Center with [e]—[æ] to show the same phonemics with different allophonics; later on we shall discuss this difference in more detail.

We may note in passing that it is possible to make the above phonemic reading without knowing which vowel quality ([e], [ɛ], [æ]) belongs with which word group (*Vetter*, *Wetter*, *Wespe*) at any given locality. It is enough to know where there is a two-way and where a three-way opposition, and what the approximate phonetics are.

For a sensible phonetic reading of Map 1 we must of course identify

[4] I am indebted to Prof. Hotzenköcherle for allowing me to use part of the base map of the Atlas. On this map, localities are numbered within each canton, and the cantons are indicated by the abbreviations familiar from auto license plates. Our area contains: 3 points in BA (Basel), 3 in SO (Solothurn), 59 in AG (Aargau), 16 in LU (Luzern), all 11 in SH (Schaffhausen), all 66 in ZH (Zürich), 7 in ZG (Zug), 6 in SZ (Schwyz), 24 in TG (Thurgau), 20 in SG (St. Gall), 3 in AP (Appenzell), and 1 in GL (Glarus). By way of orientation: AG 34 is the city of Aarau, ZH 37 the city of Zürich, SH 7 the city of Schaffhausen, TG 15 the city of Frauenfeld.

the word groups involved; the resulting pictures are then conventional dialect maps. Such a map of the *Wespe* group can be visualized by picking out the bottom symbol at each locality; we then see horizontal lines (=[æ]) throughout the West, Center, and East, but diagonal lines (=[ɛ]) in the North. This reflects the structural boundary separating the /E/—/æ/ and /e/—/ɛ/—/æ/ areas from the /e/—/ɛ/ area. A phonetic map of the *Wetter* group can be visualized by noting the middle line at each locality; we then see horizontal lines (=[æ]) in the West and Center, but diagonal lines (=[ɛ]) in the North and East. This reflects the other structural boundary, that separating the /E/—/æ/ area from the /e/—/ɛ/ and /e/—/ɛ/—/æ/ areas. A phonetic map of the *Vetter* group can be visualized by picking out the top line at each locality; we then see diagonal lines (=[ɛ]) in the West (and at SG 22, 23, 28), vertical lines (=[e]) in the Center, North, and East. A comparison of the phonemic and phonetic readings for this third group is particularly interesting. On the one hand it shows a phonetic boundary running through a phonemically uniform area: Western [fɛttər] vs. Central [fettər] in the West & Center /E/—/æ/ area. And on the other hand it shows a three-way phonemic boundary running through a phonetically uniform area: uniform [fettər] throughout the Center, North, and East, though in the Center it belongs to a structure /E/—/æ/, in the North to a structure /e/—/ɛ/, and in the East to a structure /e/—/ɛ/—/æ/.

§2. BACK VOWELS. In processing the phonetic data of Map 1 for cartographic presentation we chose three groups of five words each, rejected the clearly aberrant vowels as belonging to some other phoneme, and then averaged the values of the remaining vowels within each group. By this technique the allophonic variations of a phoneme at any one locality are leveled out, though those between two areas (Western /E/=[ɛ] vs. Central /E/=[e]) become all the clearer. By way of contrast, Map 2 is based on only three words: *Ofen* 'stove', *Hosen* 'trousers', *Wasen* 'pieces of sod'; and each line on the map represents the stressed vowel of one of these words.[5] This simpler technique has the not inconsiderable advantage of being five times as fast to prepare, since one needs to gather only 660 bits of data (3×220 localities) rather than 3,300 (15×220 localities). It also gives some idea of the allophonics at any given locality, though at the price of obscuring the phonemics.

[5] At a few points where *Wasen* was not recorded, the symbol on the map represents the vowel of *Nase* or *Wade*. At four points the symbols on the map are incorrect: AG 25, 26 should show two vertical lines, one horizontal line; AG 37, 55 should show two diagonal lines, one horizontal line.

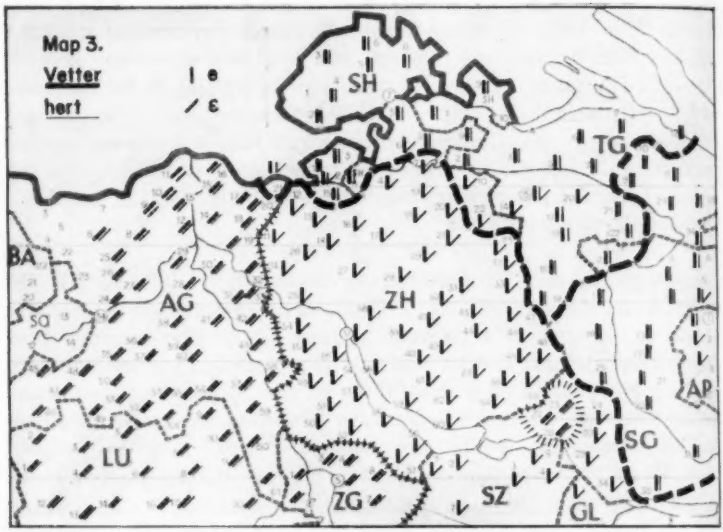
The phonemic reading of Map 2 is clear in most of the area. The North and East show a three-way contrast /o/—/ɔ/—/a/ in /ofə həsə wasə/; the isogloss drawn on the map represents the outer limits of this area. The Center and the Southwest show a two-way contrast which we might note as /O/—/a/: it is [o]—[a] ([ofə hosə wasə]) in the Center, [ɔ]—[a] ([ɔfə həsə wasə]) in the Southwest. The remainder of the West is problematical. The lack of symbols in the extreme Northwest reflects the fact that the words we have chosen show long vowels here: [ōfə hōsə wāsə]. The remainder of the area shows, in these particular recordings of these particular words, the phonetic contrasts [o]—[a], [ɔ]—[a], and [o]—[ɔ]—[a]. However, the random distribution of [o] and [ɔ] suggests that the contrast between them is not relevant. This is supported by the fact that, where [o] and [ɔ] both occur, the contrast is sometimes [ofə]—[həsə] (thus at AG 40, 56), sometimes [hosə]—[ɔfə] (thus at AG 39, 41, 42). Nevertheless, on the basis of these two words alone we could hardly settle the phonemic status of [o] and [ɔ] in this area. In particular, we could not be sure that the contrast [hosə]—[ɔfə] at AG 16 and 17 is allophonic, whereas the contrast [ofə]—[həsə] at AG 21 is phonemic. To be reasonably sure of the phonemics we need more data. Our worksheets list the stressed vowels of 17 different words: 5 which in the North and East (within the isogloss) show a consistent [o], 5 which show a consistent [ɔ], and 7 which show [o] consistently

in part of the area, [ə] consistently in the rest of it. These fuller data make it quite clear that the North and East have a three-way contrast /o/—/ə/—/a/, but that the remainder of the area has only a two-way contrast /O/—/a/. This latter is realized in the Center as [o]—[a], in the Southwest as [ə]—[a], and in the remainder of the West at some points as [o∼ə]—[a], at others as [o]—[a]. The irregularities which Map 2 shows in the West come from the fact that our system of symbolization (vertical line for [o], diagonal line for [ə]) at many points cuts right through the allophonic range of the phoneme /O/.

§3. A PROBLEM OF COMPLEMENTARY DISTRIBUTION: [e] AND [ɛ]. The material underlying maps 1 and 2 illustrates the several ways in which the phonemic relevance of a phonetic contrast can be demonstrated. Only in very rare cases do the Atlas materials offer by chance a genuine minimal pair, e.g. North and East [fešt]≠[fɛšt] '*fest, Fest*'. Usually we must accept the evidence of several pairs which show the two phones under consideration in more or less similar environments, e.g. SG 1 [fɛšt]—[wæšpəli] '*Fest, Wespe*', [rɛxxə]—[pæxtər] '*Rechen, Pächter*', [fɛrsə]—[færbə] '*Ferse, färben*'. Sometimes the Atlas materials offer only negative evidence, but this can be supported by positive evidence from individual monographs. As an example, for SG 27 our worksheets show 6 forms with close [o], 14 with open [ə]. The two phones cannot be in free variation, because they show the same distribution in other nearby localities, and Map 2 has shown us the picture of random geographical distribution given by phones which are in free variation. They cannot be in complementary distribution, unless we are willing to accept an extraordinarily complicated type of complementation. From sheer negative evidence we must therefore set them up as contrasting phonemes. We are supported in this by such positive evidence as the following pairs taken from *BSG* 9 (Toggenburg): [offə]—[trəffə] '*offen, getroffen*', [hoffə]—[ksəffə] '*hoffen, gesoffen*', [xlobə]—[kxləbə] 'thick stick, pinched', etc.

A real problem of interpretation arises when the limited materials of the Atlas suggest that two phones are in complementary distribution. We can then never be sure that further materials would not show a relevant contrast between the two, even though its functional load might be extremely small. Such an example is illustrated in Map 3, which contrasts the vowel of the *Vetter* group with that of the word *hert* 'hard'. The heavy isoglosses on this map are taken from Map 1; the shaded isogloss separates the Western [fɛttər hɛrt] area (plus SG 22, 23, 28) from the Central [fettər hɛrt] area.

If we assume that the evidence before us is complete, we are justified



in making the following phonetic and phonemic readings of Map 3. Phonically the vowel of *hert* appears as [ɛ] throughout the entire West and Center; allophonically its status in these two areas is quite different. In the West, [ɛ] is the "normal," "regular" allophone of /E/, i.e. the allophone which appears (with due allowance for free variation) in all phonological environments. In the Center, [ɛ] is a "special" allophone of /E/, i.e. one that occurs only in the environment "before /r/." This knowledge now puts the boundary between West and Center in a new light. In the West the phoneme /E/ is composed of a single acoustic target [ɛ], and deviations from this are only the usual random deviations that we call free variation. In the Center the phoneme /E/ is composed of two acoustic targets: [e] in most environments but [ɛ] before /r/, and this deviation is of the type that we call complementary distribution. The boundary is therefore structural in a very special sense of the word: it is the boundary between two areas in which a comparable phoneme has different internal, allophonic structure.

A different situation arises wherever [hɛrt] with [ɛ] has spread to, or developed independently in, the North and East, beyond the heavy isogloss on Map 3. In the Center, e.g. at ZH 20 and 21, the [ɛ] of [hɛrt] belongs structurally, as a special allophone, to the /E/ of the /fEttər/ = [fettər] group. But in the North, e.g. at the nearby villages ZH 10 and 22, the [ɛ]

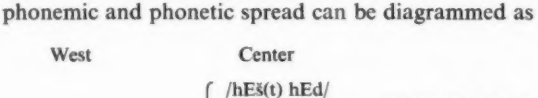
of [hɛrt] belongs structurally not to the /e/ of the /fettər/ = [fettər] group but rather, as a normal allophone, to the /ɛ/ of the /wɛttər/ = [wɛttər] and /wɛšpi/ = [wɛšpi] groups. In crossing this structural boundary the phone [ɛ] has, as it were, shifted its phonemic allegiance.

Our interpretation of [ɛ] in the Center as a special allophone of /E/ before /r/ is supported by other forms in the Atlas materials: *Herbst*, *Gertel* 'garden knife' (lacking at many points), *Kerze*, *März* (though these latter two usually show a long vowel). Further, it makes good historical sense: in the Center, MHG *e* remained a close vowel in most environments, but was lowered before a following /r/— a familiar phenomenon in German dialects. In addition, it agrees with Weber's findings in *BSG* 15 for the Zürich Highlands (the southeast section of the canton). Weber notes in §§25, 58 that MHG *e* gives [e] in most environments, but [ɛ] before /r/: [hɛrt] 'hard', [hɛrpšt] '*Herbst*', [fɛrtɪg] '*fertig*', [mɛrkxə] '*merken*', etc. He adds, to be sure (§§25, 56), a fact which the limited Atlas materials could not reveal, namely that [ɛ] also occurs before /x/: [hɛxt] '*Hecht*', [šwɛxxər] '*schwächer*', [šwɛxxə] '*schwächen*', [šwɛxxɪ] '*Schwäche*'. However, this merely gives a new environment for the special allophone. The complementation is still complete: [ɛ] occurs only before /r/ and /x/, and [e] occurs only in other environments.

The thoroughness of Weber's monograph on the Zürich Highlands establishes beyond all reasonable doubt the fact that, in this part of the Center, [e] and [ɛ] occur in complementary distribution and are therefore allophones of a single phoneme /E/. We must then ask: Is this true for the remainder of the Center? There are good structural reasons for subjecting this complementation to close scrutiny. The corresponding long vowels [ē] and [ɛ̄] are clearly in opposition throughout the Center, and therefore constitute the phonemes /ē/ and /ɛ̄/: /hēr/—/hɛ̄r/ '*Heer, her*', /dēt/—/tɛ̄t/ '*dort, täte*', /mē/—/nɛ̄/ '*mehr, nehmen*', etc. The theory of distinctive features might lead us to believe that an opposition relevant among the long vowels would spread to the short vowels as well.

A reasonably thorough search through the Atlas materials has revealed, within these materials, only six points where the complementary distribution of [e] and [ɛ] does not hold. At ZH 25, 34, 50–53 the phone [ɛ] occurs not only before /r/ (and presumably /x/) but also before /š/ and /d/ in the forms [hɛš(t)] '(*du*) *hast*' and [hɛd] '(*er*) *hat*', which we must therefore analyze as /hɛš(t)/ and /hɛd/, with a phoneme /ɛ/. The techniques of dialect geography reveal the interesting origin of this /ɛ/. In the West, *hast* and *hat* show the phoneme /E/, here realized of course as [ɛ]: /hEš(t) hEd/ = [hɛš(t) hɛd]; in the Center they show the phoneme /æ/: /hæš(t) hæd/. Historically the Western forms must have spread across the usual West/Center boundary

(the shaded isogloss of Map 3) into the southwest of the Central area. Where it was the phoneme /E/ which spread, the result was of course still /hEš(t) hEd/, but this changed phonetically from [hɛš(t) hɛd] to [heš(t) hed], with [e] as the regular allophone of /E/ in this environment. This was the solution adopted at ZH 35, 36, 49, 55. Where, on the other hand, it was the phone [ɛ] which spread, the result was of course still [hɛš(t) hɛd], but this changed phonemically from /hEš(t) hEd/ to /hɛš(t) hɛd/, with /ɛ/ now a phoneme opposed to the /e/ of /gešt/ '*gestern*' and the /æ/ of /næšt/ '*Nest*' etc. This was the solution adopted at ZH 25, 34, 50–53. This difference between phonemic and phonetic spread can be diagrammed as follows:



Outside of the Atlas materials I have found evidence for an [e]–[ɛ] opposition in the dialect of the city of Zürich. My informant for this area regularly uses [ɛ] not only before /r/ and /x/, but also before nasals: [dɛmm wɛmm dɛnə lɛŋər lɛŋšt lɛŋi xrɛŋkxər] '*dem, wem, denen, länger, längst, Länge, kränker*'.[6] If this were the whole story, it would merely add still a further environment for the special allophone [ɛ]. But there are irregularities. My informant uses close [e] in [séməlɪ] '*Semmeln*', [seméštər] '*Semester*', and [wénɪg] '*wenig*'. In addition, analogy and borrowing have both worked to introduce the phone [ɛ] into environments before consonants other than /r x m n ŋ/. On the one hand, through a complicated set of analogies (probably involving, among others, the form [dɛrɪg] 'such' with historically regular [ɛ] before /r/), there has arisen a form [sɛttɪg] 'such', whose [ɛ] before /t/ contrasts with the [e] of [bettər] '*Betten*' and

[6] Weber, *Zürichdt. Gram.*, also notes [ɛ] before /r x m n ŋ /. §41 [hɛrt šwɛxxər]; correct the statement that [ɛ] occurs only before /r/ and /x/. §127, note 3 [xrɛŋkx(n)ər]; my informant does not use the form [fɛltš(n)ər] 'falser' noted here. In §127, note 2, Weber notes [leŋər leŋšt]; my informant says [lɛŋər lɛŋšt]. §141 [dɛmm dɛnə]. §149 [wɛmm]. Of these forms with [ɛ] before nasal, only *wem* is included in the Atlas questionnaire. The Atlas materials show [wɛmm] and [wemm] distributed more or less at random in the Center, with no evidence of a contrast.

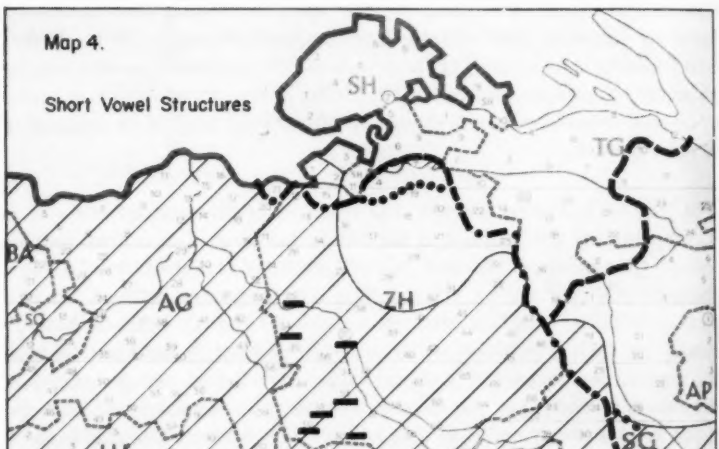
the [æ] of [wættər] '*Wetter*'. A similar historically irregular [ɛ] has arisen in the subjunctives [gɛb] '*gebe*' and [tɛg] '*tue*', probably under the influence of the long /ɛ̄/ in the conditionals [gɛ̄b] '*gäbe*' and [tɛ̄t] '*täte*'.[7] On the other hand, borrowing from standard German has introduced the use of [ɛ] in the names of several letters of the alphabet. The name [ɛr] 'R' and probably also [ɛmm ɛnn] 'M, N' fitted into the regular allophonic pattern; but the names [ɛf ɛll ɛss tsɛt] 'F, L, S, Z' did not, and hence introduced such contrasts as [ɛll] 'L' but [elə] '*Elle*' (measure of length), [tsɛt] 'Z' but [tsettə] 'to ted hay'. Since Switzerland shares in the modern liking for alphabetical names, the phone [ɛ] also appears in such well established forms as [éspebē] 'SBB' = '*Schweizerische Bundesbahnen*' = 'Swiss Federated Railways'. As a result of these various developments, an independent phoneme /ɛ/ must be posited for the dialect of the city of Zürich. It seems to be well established, even though the functional load of its opposition to /e/ is minute.

§4. FRONT ROUNDED VOWELS. The structure of the mid and low vowels is completed by the addition in the North and East of a contrasting pair /ö/—/ɔ̈/ (parallel to /e/—/ɛ/ and /o/—/ɔ/), in the West and Center of a single phoneme /Ö/ (parallel to /E/ and /O/). The boundary between the two areas is identical with that between the /o/—/ɔ/ and /O/ areas on Map 2, except that SG 35 (with /ö/—/ɔ̈/) belongs this time to the East. Like its unrounded counterpart /E/, the phoneme /Ö/ shows two geographical allophones: close [ö] in the Center (in most environments) and open [ɔ̈] in the West, entirely parallel to [e] and [ɛ]. The Atlas materials also show that, in parts of the Center, /Ö/ runs parallel to /E/ in showing a special positional allophone [ɔ̈] before /r/ and /x/: [tɔ̈rn] rather than [törn] 'thorns' appears at most points in western ZH (in eastern ZH the usual form is [tɔ̈̄rn] with a long vowel); [šɔ̈xlɪ] rather than [šöxlɪ] 'shocks of hay' appears consistently in southern ZH, inconsistently in the rest of ZH. Nowhere in the Center is there any evidence of a relevant contrast between [ö] and [ɔ̈]. A full presentation of all these phenomena would require several maps; we therefore limit ourselves to this brief verbal description.

[7] Weber, *Zürichdt. Gram.*, notes not only (§147) [sɛttig] 'such' but also (§147, note 1) archaic [sɛlig] 'such' and (§151) rural [wɛttig] 'what sort of'. The demonstrative pronoun and adjective [dɛ̄] 'he, that' (masculine) also occurs in a shortened form [dɛ] (§142), contrasting with [me] shortened from [mē] '*mehr*' (§158, p. 151, note 3). (My informant also uses these shortened forms.) It is difficult, however, to determine whether these are phonemically short vowels or prosodic allophones of long vowels, since they occur only as variants of long vowels. Weber (§303.1, 2, 3) also notes [ɛ] in the interjections [hɛ ɛhé sɛ].

§5. SYSTEMS OF MID AND LOW VOWELS. By combining the isoglosses of maps 1, 2, and 3 we are now able to indicate in Map 4 the systems of mid and low vowels occurring within our area.[8] We find the following:

West & Center	Special ZH	North	East	ZH 4, 7	SG 35
E Ö O	e Ö O ɛ	e ö o ɛ ɔ̈ ɔ	e ö o ɛ ɔ̈ ɔ	e ö o ɔ̈ ɔ	e ö o ɛ ɔ̈
æ a	æ a	a	æ a	æ a	æ a



The systems labeled "West & Center" and "Special ZH" occur in the following approximate phonetic shapes:

Southwest	Remaining West	SG 22, 23, 28	Center	Special ZH
ɛ ɔ̈ ɔ	ɛ ɔ̈ {o / ɔ}	ɛ ɔ̈ o	[e / ɛ] [ö / ɔ̈] o	e [ö / ɔ̈] o ɛ
æ a	æ a	æ a	æ a	æ a

[8] Points ZH 25, 34, 37 (city of Zürich), 50–53 are underlined to indicate the presence here of a phoneme /ɛ/, giving the front vowel structure /e/—/ɛ/—/æ. This is referred to below as the "Special ZH" system; it may perhaps occur also in other parts of the Center. The shading on Map 4 will be explained in §6 below.

Allophones in free variation are linked by braces, allophones in complementary distribution are linked by square brackets. Close [e] and [ö] also occur as free variants in the area labeled "Remaining West," but they are so much rarer than close [o] that they are not included in the above diagram.

§6. HIGH VOWELS. The short vowel systems of our area also contain high vowels, either the single set /i ü u/ or the opposed sets /i ü u/≠/ɪ Ü ʊ/. To oversimplify, the single set /i ü u/ occurs in the Northeast and the opposed sets /i ü u/≠/ɪ Ü ʊ/ in the West and Southeast, whereas in the Center there lies an area in which the opposition /i ü u/≠/ɪ Ü ʊ/ can be neither fully supported nor fully denied. One would like to find relatively clear boundaries between these areas. We have collected a large amount of material in a search for such boundaries, only to come to the conclusion that they do not exist. This in itself is a useful finding. Synchronic phonemics assumes that phonemes are discrete, non-overlapping elements, and that the transitions from one phonemic system to another must therefore be completely sharp. Diachronic phonemics, on the other hand, assumes that during the course of time one phoneme can split into two, or two coalesce into one, and every such change implies a period of phonemic indeterminacy in part of the system. Dialect phonemics, it seems, must make both assumptions. On the one hand it must seek a synchronic analysis which will reveal such completely sharp transitions from one system to another as those which we have seen thus far. And on the other hand, like diachronic phonemics, it must allow for gradual transitions with resulting phonemic indeterminacies, such as the one we are about to describe.

First, some synchronic data. The following pairs show identical vowels in the canton of Schaffhausen (data from *BSG* 20) but contrasting vowels in the Zürich Highlands (data from *BSG* 15):

SH /i ü u/		ZH Highlands /i ü u/≠/ɪ Ü ʊ/		
šwigə	= rigəl	šwigə	≠ rɪgəl	'*schweigen, Riegel*'
gitslə	= sitsə	gitslə	≠ sɪtsə	'to bear young (of goats), *sitzen*'
æksküsi	= xüssi	æksküsɪ	≠ xÜssɪ	'excuse me, *Kissen*'
gükkəl	= tsügəl	gükkəl	≠ tsügəl	'rooster, *Zügel*'
gukkə	= rukkə	gukkə	≠ rʊkkə	'*gucken, Rücken*'
buslə	= dussə	buslə	≠ dʊssə	'pussy, *draussen*'

Next, some historical interpretation. The second word in each pair above represents the normal reflex of MHG short /i ü u/: MHG *rigel, sitzen, küssin* (standard German *Kissen* shows unrounding), *zügel*, and *rücke* (the dialect forms lack umlaut). SH *dussə*, ZH *dʊssə* is from MHG *dā ūzen* 'there outside', with early shortening of /ū/ to /u/, which then coalesced

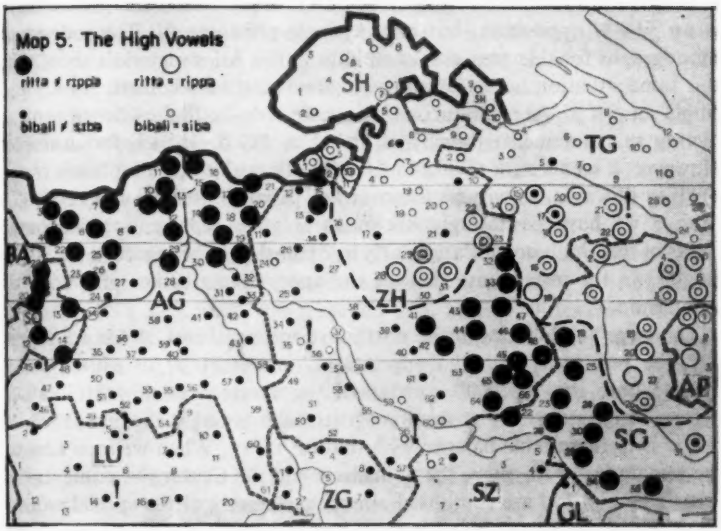
with the usual short /u/. Of the first words in each pair, the only one with a reasonably clear history is *šwigə* < MHG *swīgen*. Weber (*BSG* 15, §65.1) gives the generally accepted explanation for the shortening of the vowel: "Kürzung tritt ein unter dem Einfluss des stark geschnittenen Silbenakzentes bei starkem Expirationsdruck," for which statement he gives several examples, among them the imperative (MHG *swīg* >) *šwig* 'be quiet!', whence analogically the whole present tense took on the shortened vowel. SH *ɛkskúsi*, ZH *ækskúsɪ* is of course a loan, from French *excusez*. The remaining forms all seem to be ultimately of imitative origin. The call to goats is *gits gits*, whence SH *gitsli*, ZH *gitsɪ* 'kid' and the verb *gitslə*. 'Cock-a-doodle-do' is ZH *gükkərükkü*, whence *gükkəl* 'rooster'. The call used in playing hide-and-go-seek is ZH *gukkūs*, whence *gukkə* 'to look, (or perhaps *gukkə* is to be reconstructed as **guggjōn*, with the call derived from this and its imitative close vowel then transferred back to the verb). The call to cats is ZH *bús bús*, whence *buslə* 'pussy'.

If the only source of close /i ü u/ were these "special" words (as we shall call them), then the opposition /i ü u/ ≠ /ɪ ʊ̈ ʊ/ (where it occurs) would be extremely unbalanced, since there are very few special words as against very many words with open /ɪ ʊ̈ ʊ/ < MHG /i ü u/. In the ZH Highlands, however, the incidence of close /i ü u/ was greatly increased by a regular shortening of long /ī ǖ ū/ before fortis stops:[9] *tsit*, *xrüts*, *brut* < MHG *zīt*, *kriuze* (*iu* = /ǖ/), *brūt* '*Zeit*, *Kreuz*, *Braut*', and an irregular shortening before medial lenis consonants: *sidə*, *isə*, *šuflə* < MHG *sīde*, *īsen*, *schūvel* '*Seide*, *Eisen*, *Schaufel*'. As a result, the incidence of /i ü u/ is in reasonable balance with that of /ɪ ʊ̈ ʊ/.

To plot the areas where the opposition /i ü u/ ≠ /ɪ ʊ̈ ʊ/ does and does not occur, we have used a double procedure. First, from the Atlas materials we selected a "special" word which shows a (more or less) consistently close high vowel: *bíbəli* 'chicken, chickie', derived from *bíbíbí*, the call to chickens; and we compared its stressed vowel with that of *sieben* < MHG *siben* 'seven'. Secondly, we selected four words with MHG long /ī/, shortened before fortis stop in parts of our area to /i/: *reiten*, *Scheit*, *Freitag*, *weit* < MHG *rīten*, *schīt*, *vrītac*, *wīt*; and we compared their stressed vowels with those of four words containing MHG short /i/: *Rippe*, *flicken*, *Genick*, *Schlitten* < MHG *rippe*, *vlicken*, *genicke*, *slite*.[10]

[9] Undisturbed phonetic change would have neutralized the short ≠ long opposition before fortis stops, with only short /i ü u/ occurring in this position. However, analogy restored the opposition: *līt* '(*er*) *liegt*' (probably after *līšt* '(*du*) *liegst*'), *šrīpšt*, *šrīpt* '(*du*) *schreibst*, (*er*) *schreibt*' (after infinitive *šrībə* etc.).

[10] The evidence of *Schlitten* was discounted in parts of TG, SG, and AP, where by a special development (see below, §7) it appears with the vowel /e/: *šletə*, as against /i/ in the other examples of MHG /i/.



The result is Map 5. Large circles represent the points where MHG /ī/ was shortened before fortis stops. If this shortened vowel has coalesced with the reflex of MHG /i/, the circle is left empty: /rittə/ = /rippə/ '*reiten, Rippe*'. If the two are in opposition, the circle is filled: /rittə/ ≠ /rɪppə/.[11] Small circles represent the points where the words *bíbəli* and *sieben* occur. If the stressed vowels of the two words are the same, the circle is left empty: /bíbəli/ = /sibə/. If they are different, the circle is filled: /bíbəlɪ/ ≠ /sɪbə/.[12]

[11] *Rippe* also appears as /rippi rip/ or /rɪppɪ rɪp/.

[12] *bíbəli* also appears at some points as /bíbi/ or /bíbī/, and at SG 23 and 30 as /didəlɪ/; *sieben* also appears as /sibni/ or /sɪbnɪ/.

Because the Atlas workers' transcription is so detailed, an arbitrary grouping of their notations into "same" and "different" had to be made. We classed as "close" all the vowels which they noted as over-close, close, or neutral; and we classed as "open" all the vowels which they noted as open, over-open, or between [i] and [e]. On Map 5 the marking "same" (empty circle) then means that both vowels are "close": /rittə/ = /rippə/, /bibəli/ = /sibə/; the marking "different" (solid circle) means that the stressed vowels of /rittə bibəli/ are "close," those of /rɪppə sɪbə/ "open". To this general statement there are six exceptions, each indicated by an exclamation point on the map. At LU 5, 7, 14 the Atlas materials note the vowel of *bibəli* as open, that of *sieben* as over-open; I nevertheless classed them as "different." At ZH 35, 56 the Atlas materials note the vowels of both *bibəli* and *sieben* as open; they are therefore "same," though not in the usual way. At TG 21 the Atlas materials note the vowel of *bibəli* as open, that of *sieben* as close, i.e. the reverse of the usual relationship; I interpret both qualities as lying within the allophonic range of the phoneme /i/, and therefore marked them as "same."

Map 5 shows in the Northeast and East a considerable area where there is no /i/≠/ɪ/ opposition, but only a single phoneme /i/. The individual monographs for this area plus samplings in the Atlas materials show that this same situation holds for all the high vowels: they constitute only the single set /i ü u/. At a number of points we seem to find evidence contradicting this statement: [bibəli] vs. [sɪbə] at TG 5, 15, 17, for example. However, it seems clear that both [i] and [ɪ] lie within the allophonic range of /i/ in this area, as witness the contrary [sibə] vs. [bíbəli] at TG 21. On Map 5 we have hesitatingly drawn an isogloss representing the outer limits of this /i ü u/ area. Particularly in ZH it should be regarded less as an isogloss in the usual sense than as the approximate center of a zone of indeterminacy.

The large solid circles in the northwest and southeast of Map 5 represent the two areas with the opposition /i ü u/≠/ɪ Ü ʊ/ in which /i ü u/ include both the "special" words and the shortenings of MHG /ī ǖ ū/ before fortis stops. As a result of these shortenings, the incidence of /i ü u/ is in reasonable balance with that of /ɪ Ü ʊ/. When we turn now to the remainder of the map, the Southwest and the Center, we come to the areas where /i ü u/ are composed almost exclusively of the special words. Here the incidence of /i ü u/ is therefore quite out of balance with that of /ɪ Ü ʊ/.

What are the structural implications of this imbalance? The crucial factor seems to be the phonetic interval between the realizations of the two sets of phonemes. In the Southwest, /i ü u/ are generally realized as close vowels and /ɪ Ü ʊ/ as over-open vowels, ranging even as open as [e ö o]; Map 5 shows that in this area the opposition /i ü u/≠/ɪ Ü ʊ/ is consistently maintained. In the Center, /i ü u/ are realized as close vowels and /ɪ Ü ʊ/ as neutral or open, but seldom beyond that; Map 5 shows that in this area the opposition /i ü u/≠/ɪ Ü ʊ/ is *not* consistently maintained. If we plot the whole short vowel structures of these areas, it becomes clear that the size of this interval is a function of the other phonetic intervals in the systems:

Southwest			Center			ZH Highlands			North		
i	ü	u	i	ü	u	i	ü	u	i	ü	u
ɪ	Ü	ʊ	ɪ	Ü	ʊ	ɪ	Ü	ʊ			
			[e ɛ]	[ö ɔ̈]	O	[e ɛ]	[ö ɔ̈]	O	e	ö	o
E	Ö	O							ɛ	ɔ̈	ɔ
æ		a	æ		a	æ		a		a	

Except in the Southwest, the interval between /i ü u/ and /ɪ Ü ʊ/ is the smallest within the system. Where, as in the ZH Highlands, the incidence

of /i ü u/ is reinforced by shortenings from MHG /ī ǖ ū/, the opposition is in reasonable balance and seems to be clearly maintained. Where, as in the Center, the incidence of /i ü u/ is limited to the "special" words, the opposition is in imbalance and is not clearly maintained.

The city of Zürich (ZH 37) serves as an excellent example of an area in which the opposition /i ü u/ ≠ /ɪ ʏ ʊ/ is in imbalance and is not clearly maintained. I choose it because of the unusually full data I have on it: the Atlas records of three different elderly informants A, B, and C, and my own extensive records of a university student informant. In a random sampling of 40 words representing MHG /i ü u/, the Atlas informants show the following ratios of open to close: A, 40:0 (i.e. all open); B, 32:8; C, 16:22. (C's 39th word was recorded with both qualities, his 40th was not recorded.) Some examples: A [xɪlə], B [xɪlə], C [xilə] '*Kirche*'; A [tütšə], B [tütšə], C [tütšə] 'to strike'; A [rʊkkə], B [rukkə], C [rukkə] '*Rücken*'. Particularly interesting are the following two words with shortened MHG /ī/: A [xridə sɪdə], B [xrɪdə sɪdə], C [xrɪdə sidə] '*Kreide, Seide*'.[13]

In the case of my own Zürich informant, I found only a handful of words which seem to show a consistently close vowel. They fall into two categories: (1) imitative words, such as [bíbi]˙'chicken', [gígələ] 'to giggle', /búspús/ 'here pussy'; and (2) loans, such as [profít] 'profit', [wisíttəxə̀rtə] '*Visitenkarte*', [ekskʊ́sɪ] 'excuse me', [tüp] 'type'. The vast majority of words, whatever their origin, appear with a quality that varies freely over the whole range from open to close, with a more or less neutral quality as the commonest. Only in a few cases do my records show the notation "never fully close": [šnɪt], not [šnit] '*Schnitt*'.

In addition to *bíbəli* 'chicken', which I have incorporated into Map 5, the Atlas materials show three other "special" words whose vowels are of interest:

(1) *gitsi* 'kid' and *gitslə* 'bear young (of goats)' show close [i] in most of the area, but open or over-open [ɪ] = /ɪ/ consistently in the Northwest, the Southwest, and in ZH along and southwest of the lake.

(2) *gükkəl* 'rooster' shows close [ü] quite consistently in SH, TG, and most of ZH. However, it has open or over-open [ʏ] = /ʏ/ in AG, LU, ZG, SZ and in ZH along and southwest of the lake. SG (except for 22, 23, 28

[13] The shortening of MHG /ī ǖ ū/ in open syllable before lenis, mentioned above for the ZH Highlands, appears throughout ZH in a few words. Some of the above differences between A, B, and C undoubtedly come from the fact that one Atlas worker made the records for A and B, another for C; but note that neither A, B, nor C shows complete historical consistency. The few "special" words in the Atlas materials cannot be illustrated for the city of Zürich because they all occur in farming sections of the questionnaire which were not asked in big cities.

with open [ü] and 24, 29, 30 with close [ü]) shows a different form of the word: *gugəl*, usually with close [u].

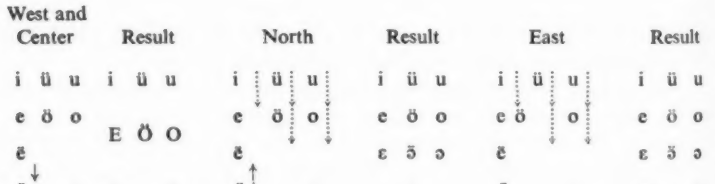
(3) *súksúk* 'call to pigs' shows close [u] fairly consistently in most of ZH and TG, but otherwise it is a poor example. Most of AG uses *hæshæs* and the like, and the south of our area has *hʊrshʊrš* and *hóshós* beside *húshús*.

In order to complete the short vowel picture of Map 4, we have added to it the isogloss of Map 5. Shading indicates the opposed sets /i ü u/ ≠ /ɪ ʏ ʊ/, absence of shading the single set /i ü u/. It should be emphasized again, however, that this "isogloss" is better interpreted as the probable center of a zone of phonemic indeterminacy.

§7. HISTORY. The history of these short vowel systems illustrates a number of interesting points. With the exception of the vowels of the "special" words and the shortenings of MHG /ī ǖ ū/, the systems can all be derived from the one that is customarily assumed for MHG:

i	ü	u
e	ö	o
ë		( )
ä		a

We give this system with parentheses around what is an obvious hole in the pattern. This asymmetry was removed in our area in the following three ways (solid arrows refer to phonemic coalescences, broken arrows to phonemic splits):



The asymmetry of the MHG system lay in the fact that the front vowels contained one more relevant level than the back vowels. In the West and Center this asymmetry was removed by decreasing the number of front

vowels: *ë* (*Wetter*) was lowered until it coalesced with *ä* (*Wespe*), giving /wættər/ = /wæšpi/.[14] In the North and East the asymmetry was removed by increasing the number of back vowels: the /o/ of MHG *ofen, hose* split into modern /ofə/ ≠ /həsə/.[15] The parallel split of /ö/ into /ö/ ≠ /ɔ̈/ was not required for reasons of symmetry, since the central column of such a three-column system nearly always contains fewer levels than do the front and back columns.[16] Hence morphologically isolated words such as *plötzlich* kept their close /ö/ (even, as here, before dental obstruents). However, where MHG /ö/ alternated morphophonemically with /o/ (as in sg. *boc*, pl. *böcke*, '*Bock*, *Böcke*'), analogy led to the development of a new phoneme /ɔ̈/: pl. /bökx/ in analogy with sg. /bəkx/. (Part of the area still preserves analogically undisturbed alternations of the type /bəkx bökx/.) This widespread phonetic and analogical lowering of /ö o/ to /ɔ̈ ə/ would have led to a new type of asymmetry: the incidence of retained close /ö o/ would have been extremely small and quite out of balance with the incidence of the other short vowel phonemes. Such an asymmetry seems never to have arisen, however, because the incidence of /ö o/ was increased by a split of the reflexes of MHG /i ü u/ into /i ü u/ ≠ /e ö o/: retention as /i ü u/ for example in /sibə xütslə šuppə/ '*sieben, kitzeln*, tuft of hair', but lowering to /e ö o/ in /kšrebə šössle soppə/ '*geschrieben, Schüssel, Suppe*'.[17]

[14] In parts of the canton of Glarus (though not at the one point on our map) the number of front vowels was decreased by a change in the opposite direction: *ë* (*Wetter*) was raised until it coalesced with *e* (*Vetter*), giving /wettər/ = /fettər/:

i	ü	u	Result:	i	ü	u
e	ö	o		e	ö	o
ë ↑						
ä		a		æ		a

The resulting structure is identical with that labeled "West and Center" above. The difference is one of incidence: /wettər/ = /fettər/ rather than /wættər/ = /wæšpi/. Cf. Catharina Streiff, *Die Laute der Glarner Mundarten*, §§24, 30 (1915; = *BSG* 8).

[15] The start of this split seems to have been the development of a lower allophone of /o/ in certain environments, especially before dental obstruents and /r/: [hosə] ~ [həsə] etc. The split became phonemic partly through various analogical developments, partly through irregular geographical spread of the new [ə], and partly through the reintroduction of [o] into environments where it had previously been lowered, as when *[wolt] '*(ich, er) will*' (with close [o] regularly retained before /l/) became [wot], contrasting with [gət] < earlier [got] '*Gott*' etc.

[16] Cf. André Martinet, *Économie des changements phonétiques*, §§3.30, 4.6 (Berne, 1955).

[17] Forms from *BSG* 1. Since the various lowerings described here affected no less than five different phonemes, /ö o i ü u/, one might be inclined to view the whole process as a

It was stated above that symmetry was attained in the West and Center through the coalescence of MHG *ë* and *ä*, resulting in a system in which the front vowel and back vowel columns show the same number of distinctive levels. When we now look at the development of this same MHG system in the North, we find that here also *ë* and *ä* coalesced. The question then arises: Why did the same coalescence produce symmetry in the West and Center but not in the North? The answer lies clearly in the phonetic nature of the phoneme which resulted from the coalescence. In the West and Center the result was low front [æ]: [wættər] = [wæšpi], where [æ] shared with [a] the distinctive feature "low," thus producing a symmetrical rectangular vowel system. In the North, on the other hand, the result of the coalescence was lower mid [ɛ]: [wɛttər] = [wɛšpi], where [ɛ] shared no distinctive feature with [a], thus producing a triangular vowel system which, if no further change had occurred, would have remained asymmetrical. Symmetry was attained through the split of /o/ into /o/ ≠ /ɔ/, just as in the East where MHG *ë* and *ä* remained distinct phonemes: /wɛttər/ ≠ /wæšpi/. This parallelism between North and East is our justification for believing that a structural analysis into rectangular and triangular systems, based on the presence or absence of low [æ] beside low [a], represents linguistic reality and is not mere analyst's play. If the diagrams which we give were merely playful manipulations of symbols on paper, then either of the following arrangements could represent the system resulting in the North from the coalescence of *ë* and *ä* as [ɛ]:

West and Center			North (1)			North (2)			East		
i	ü	u	i	ü	u	i	ü	u	i	ü	u
E	Ö	O	e	ö	o	e	ö	o	e	ö	o
æ		a	ɛ		a	ɛ		( )	ɛ		( )
							a		æ		a

The fact that the subsequent development of the vowel system of the North was parallel not to that of the West and Center but to that of the East confirms the belief that arrangement "North (2)" represents linguistic reality more faithfully, and suggests that arrangement "North (1)" would indeed be only a playful manipulation of symbols on paper.

---

general lowering of vowel articulation (such as we find, for example, in LU and southern AG). This cannot be the case, however, since (1) the phenomenon is structurally not a general lowering but a series of phonemic splits, and (2) the phoneme /e/, which was not subject to any structural pressures, remained unaffected. For a full discussion of these matters, see my article, "Lautwandel durch innere Kausalität: Die ostschweizerische Vokalspaltung," to appear in the *Zeitschrift für Mundartforschung*.

§8. COMPARING PHONEMIC SYSTEMS. The above historical remarks allow us now to diagram a few sample modern short vowel systems and to add to the diagrams an indication of the vowels' historical sources; this at the same time gives an indication of the lexical correspondences from one system to another.[18] We number the MHG vowels according to levels, from top to bottom, and then carry these numberings over into the modern systems. Subscript $/_0/$ refers to the vowels of the "special" words; subscript $/_{11}/$ refers to the shortenings of MHG long /ī ū ū/.

MHG	North	East	SG 19, 25,
i_1 $ü_1$ u_1	$i_{0,1}$ $ü_{0,1}$ $u_{0,1}$	$i_{0,1(11)}$ $ü_{0,1(11)}$ $u_{0,1(11)}$	$i_{0,11}$ $ü_{0,11}$ $u_{0,11}$ $ɪ_1$ $ʏ_1$ $ʊ_1$
e_2 $ö_2$ o_2	$e_{1,2}$ $ö_{1,2}$ $o_{1,2}$	$e_{1,2}$ $ö_{1,2}$ $o_{1,2}$	$e_{1,2}$ $ö_{1,2}$ $o_{1,2}$
$ë_3$	$ɛ_{3,4}$ $ɔ̈_2$ $ɔ_2$	$ɛ_3$ $ɔ̈_2$ $ɔ_2$	$ɛ_3$ $ɔ̈_2$ $ɔ_2$
$ä_4$ a_4	a_4	$æ_4$ a_4	$æ_4$ a_4

Southwest	Special ZH	ZH Highlands	ZH 16–21, 23, 28–31
i_0 $ü_0$ u_0	i_0 $ü_0$ u_0 $ɪ_1$ $ʏ_1$ $ʊ_1$	$i_{0,11}$ $ü_{0,11}$ $u_{0,11}$ $ɪ_1$ $ʏ_1$ $ʊ_1$	$i_{0,1}$ $ü_{0,1}$ $u_{0,1}$
$ɪ_1$ $ʏ_1$ $ʊ_1$	e_2 [$ö_2$] o_2	[E_2] [$ö_2$] O_2	[E_2] [$ö_2$] O_2
E_2 $Ö_2$ O_2	$ɛ_2$		
$æ_{3,4}$ a_4	$æ_{3,4}$ a_4	$æ_{3,4}$ a_4	$æ_{3,4}$ a_4

Such diagrams, though elaborate, have a double value: looked at diachronically, they enable us to compare the evolution of the various systems; looked at synchronically, they enable us to compare both the structure of the systems and the lexical correspondences between them.

The whole problem of comparing phonemic systems is an interesting and complicated one. In his article, "Is a Structural Dialectology Possible?"

[18] Historical sources and lexical correspondences are of course two aspects of the same thing. We reconstruct the historical sources on the basis of the modern lexical correspondences; hence the modern lexical correspondences can be inferred from an indication of the historical sources. (The vowel system which we reconstruct for our dialects is identical with that of recorded MHG, with some minor differences of phonemic incidence only. We label it "MHG" because there seems to be no good reason for preferring to call it "Proto-Swiss" or some such thing.) In indicating the historical sources, we give only the broad lines of development, ignoring such limited changes as for example MHG /e/ > ZH /æ/ before nasals: *brennen* > /brænə/ '*brennen*'.

(*Word* X [1954], 388–400), Uriel Weinreich has suggested comparison by means of a diasystem which "can be constructed by the linguistic analyst out of any two systems which have partial similarities . . ." (p. 390). His suggestions have been elaborated by G. R. Cochrane, "The Australian English Vowels as a Diasystem" (*Word* XV [1959], 69–88), who has added the very useful distinction between phonemic (diaphonemic) and subphonemic (diaphonic) comparison.

In the earlier of these two articles, Weinreich discusses briefly the problem of lexical correspondences between two or more varieties of a diasystem, but does not state whether or not they should be taken into account in setting up the diasystem. Cochrane treats the problem fully and gives a clear answer: the diasystem is set up without regard to lexical correspondences; these are handled later on. Unfortunately the problem is not as easily solved as this. If in constructing a diasystem for two or more varieties of a language we disregard lexical correspondences, then we are treating these varieties as if they were totally unrelated to one another. If, on the other hand, we treat them as related and hence take lexical correspondences into consideration, then the usefulness of the diasystem in dialectology becomes questionable. To illustrate the dilemma, we give below two diasystem analyses of the dialects of the portions of LU(zern) and AP(penzell) which appear on our map. (Except for the symbols used for LU, and the lack of /i_{11} $ü_{11}$ u_{11}/ in AP, they are the systems given above with the labels "Southwest" and "East," respectively.) The first analysis disregards lexical correspondences, the second takes them into account. In order to avoid any possible inaccuracies which might arise from our interpretation of the Atlas materials, we base the analyses on two published monographs, those of Fischer for LU and of Vetsch for AP. (See above, note 2.) Note that Fischer writes our /ɪ ʏ ʊ/, /E Ö O/ as /e ö o/, /ɛ ɔ̈ ɔ/, respectively. We have changed Vetsch's symbol /ä/ to /æ/, in agreement with our own and Fischer's practice.

LU, AP Diasystem if Lexical Correspondences are Disregarded:

LU, AP //i ≈ e ≈ ɛ ≈ æ ≈ a ≈ ɔ ≈ o ≈ u ≈ ü ≈ ö ≈ ɔ̈//

LU, AP Diasystem if Lexical Correspondences are Included:[19]

$$\textit{LU,AP}\ //\ \frac{\text{LU/i}_0\sim e_1\sim\varepsilon_2\sim æ_{3,4}/}{\text{AP/i}_{0,1}\sim e_{1,2}\sim\varepsilon_3\sim æ_4/} \approx a_4 \approx \frac{\text{LU/}ɔ_2\sim o_1\sim u_0\sim ü_0\sim ö_1\sim ɔ̈_2/}{\text{AP/}ɔ_2\sim o_{1,2}\sim u_{0,1}\sim ü_{0,1}\sim ö_{1,2}\sim ɔ̈_2/}\ //$$

[19] Since /$ɔ_2$/ and /$ɔ̈_2$/ are listed here for both LU and AP, these would seem at first glance to constitute the fully shared diaphonemes //$ɔ_2$ $ɔ̈_2$//. This is not the case, however. Whereas all the lexical items identified by subscript /$_2$/ belong to LU/$ɔ_2$ $ɔ̈_2$/, only part of them belong to AP/$ɔ_2$ $ɔ̈_2$/, the rest belonging to AP/$o_{1,2}$ $ö_{1,2}$/.

If we disregard lexical correspondences, it turns out that LU and AP share not a disasystem but the same identical phonemic system. A Luzerner and an Appenzeller will hardly agree to this. Nor will a dialectologist, since we are treating these two dialects as if they were unrelated. Yet if we treat them as related, and hence take lexical correspondences into consideration, we get a diasystem with just one fully shared diaphoneme: //a_4//. Since we are dealing here with two dialects hardly 50 miles apart, this makes the concept of the diasystem seem of questionable value in dialectology. It remains of course a highly useful concept where we wish to compare the phonemic systems of two languages which are unrelated, or whose relationship does not concern us.[20] But this is no longer dialectology.

§9. CONCLUSIONS. The exposition in §§1–8 allows us to isolate a few relatively new concepts which may prove useful in structural dialectology:

(1) GEOGRAPHICAL COMPLEMENTATION. The term "complementary distribution" customarily refers to a distribution that is phonologically (linguistically) determined. It is convenient to expand this concept to a distribution that is geographically (non-linguistically) determined. Map 1 has shown us an example of such geographical complementation: the West and the Center (excluding the "Special ZH" points) show the same phonological structure and share the same phoneme /E/; but allophonically this /E/ appears in the West as open [ɛ], in the Center as (in most environments) close [e]. Beside these "geographical allophones" we have also seen positional allophones (in phonological complementation) in Map 3: Central /E/ = [e] in [fettər] but [ɛ] in [hɛrt]; and we have seen free allophones (in free variation) in Map 2: Western [o] ~ [ɔ] in /Ofə hOsə/. These three types of distribution give quite different pictures when presented in map form.

(2) ALLOPHONIC RANGE. A matter of some structural importance is the amount of phonetic space covered by the allophones of a phoneme. We may recall that Western [hɛš(t) hɛd] spread into parts of the Center (1) phonetically as [hɛš(t) hɛd] and (2) phonemically as [heš(t) hed]. If the Center had no phone [ɛ], then spread (2) would be a simple matter of sound substitution. But since the Center already has the phone [ɛ] within the allophonic range of its phoneme /E/, the change is a more subtle one:

[20] Cf. the pioneering work done by Einar Haugen, "Problems of bilingual description," *Georgetown University Monograph Series on Languages and Linguistics* VII (1954). 9–19; "Problems of bilingual description," *General Linguistics* I (1955), 1–9; "The phoneme in bilingual description," *Language Learning* VII (1956/57), iii–iv, 17–23.

it is the readjustment of a phone to fit internal phonemic structure. It is this difference which distinguishes it from spread (1), the borrowing without readjustment of a phone already possessed. Another illustration of the effect of allophonic range is the vowel of [hɛrt] (Map 3). In the Center it belongs to /E/, the phoneme of the /fEttər/ group, which ranges here from [ɛ] to [e]. In the North and East, however, this same phonetic range [ɛ] to [e] yields the two phonemes /ɛ/ and /e/. In these areas the vowel of [hɛrt] therefore belongs no longer to the /e/ of the /fettər/ group but to the /ɛ/ of the /wɛttər/ group.[21]

(3) PHONETIC INTERVAL. Related to the concept of the allophonic range of a phoneme is that of the phonetic interval between the ranges of two phonemes. We may recall that both in parts of the Center and in the Southwest the row /i ü u/ is made up only of the "special" words; and that whereas the Center tends to confuse this row and /ɪ Ü ʊ/, the Southwest keeps them quite separate. (See Map 5.) The reason seems to lie in the different phonetic interval between /i ü u/ and /ɪ Ü ʊ/ in these two areas: in the Center it is that between close [i ü u] and open [ɪ Ü ʊ]; in the Southwest it is that between close [i ü u] and over-open [ɪ Ü ʊ] or even [e ö o]. As we have seen, in his monograph Fischer chose the latter set of symbols.[22]

(4) INCIDENCE OF A PHONEME.[23] Though traditional phonemics pays little attention to this concept, it turns out to be of great importance in dialect phonemics (and probably in general phonemics as well). We shall discuss its structural effects in item (6) below, and deal here with other aspects of the concept.

We have seen above that, as regards the phonemic inventory and the structure of distinctive features, the short vowel systems of LU and AP are identical. Three groups of observers would nevertheless object to

[21] As used here, "allophonic range" is not necessarily identical with Martinet's very useful concept of "champ de dispersion" (*Economie*, §2.10). The Central phoneme /E/ has an allophonic range which includes the positional allophone [e] of [fettər] and the positional allophone [ɛ] of [hɛrt]. Each of these positional allophones, in turn, has its own "champ de dispersion," i.e. a clustering around the quality [e] in [fettər], a clustering around the quality [ɛ] in [hɛrt] etc. As I understand it, at any rate, "champ de dispersion" refers to any dispersion around a single phonetic center (any "shots at a single acoustic target"), whether the center be that of the only allophone of a phoneme or that of one of two or more positional allophones. The term "allophonic range," on the other hand, is intended to cover the total phonetic range within which a phoneme is realized, whether it be that of a single allophone or that of two or more positional allophones.

[22] The concept "phonetic interval" is of course identical with Martinet's "marge de sécurité" (*Economie*, §2.11). Perhaps I am wrong in preferring a less vivid term.

[23] I owe this term to Hans Kurath.

calling them "the same." First, the linguistic historian will insist that only one of the eleven matching vowels in the two systems has the same history. Secondly, the dialectologist will insist that the lexical correspondences between matching vowels in the two systems are in only one case identical. (Put another way, the "lexical membership" of the phonemes is different in all but one case.) And finally, the speakers themselves will insist that the two dialects "don't sound the same" for such reasons as the fact that "in LU you say [wætter], in AP [wɛttər]." All of these objections are different ways of saying that the incidence of the corresponding phonemes in the two systems is different.

Concern with the attitudes of the speakers may seem to belong to a different level of dialect study, but occasionally these attitudes can have structural consequences. Speakers in the ZH Lake Area and speakers in the ZH Highlands are highly conscious of the difference between one another's speech. This seems at first glance to be odd, since both areas show the same identical vowel system:

i	ü	u	ī	ū	ū	iə	üə	uə
ɪ	ʏ	ʊ	ɪ̄	ʏ̄	ʊ̄			
[e]	[ö]	o	ē	ō	ō	ei	öi	ou
[ɛ]	[ɔ̈]		ɛ̄	ɔ̄			ɔ̈i	
æ		a	ǣ		ā	ai		au

As we have already seen, the Lake Area tends to confuse /i ü u/ with /ɪ ʏ ʊ/, and also /ī ū ū/ with /ɪ̄ ʏ̄ ʊ̄/. Yet this structural difference apparently goes unnoticed by the speakers themselves. The striking difference to them—a difference that has led to all sorts of rhymes and sayings making fun of one another—is the fact that many words with /ā/ in the Lake Area show /ō/ in the Highlands: Lake Area /dā gā hār šāff/ '*da, gehen, Haar, Schaf*' but Highlands /dō gō hōr šōff/ etc. (Historically, MHG lengthened /a/ gives /ā/ everywhere: *gar* > /gār/ '*gar*'; MHG /ō/ gives /ō/ everywhere: *ōre* > /ōr/ '*Ohr*'; but MHG /ā/ gives /ā/ in the Lake Area, /ō/ in the Highlands: *jār* > /jār/ vs. /jōr/ '*Jahr*'.) The great awareness of this difference in phonemic incidence has led at many points to the creation of a "compromise vowel" /ō/: *gar* always with /ā/, *Ohr* always with /ō/, but *Jahr* either /jōr/ beside /jār/, or /jōr/ beside /jōr/, or sometimes all three: /jār jōr jōr/. It is amusing, and at the same time significant, that at ZH 29 the villagers divide themselves into /ā/ and /ō/ families.

One further aspect of phonemic incidence needs to be mentioned. The discussion in §§1–7 has shown that, although the boundaries between

different phonemic structures can sometimes be very vague (as is the case with /i ü u/ ≠ /ɪ Ü ʊ/), they are more often completely sharp: a given area either does or does not have a given phoneme. The boundaries between differences in phonemic incidence within the same structure, on the other hand, seem never to be sharp. For example, the boundary between the West and Center with /O/ and the North and East with /o/ ≠ /ɔ/ is completely sharp; but the incidence of /o/ vs. /ɔ/ in the North and East is extraordinarily irregular.

(5) ALTERNATING PHONEME. The "compromise vowel" just mentioned is a familiar concept in dialect geography, and it needs to be introduced in one form or another into general phonemic theory. If such a vowel becomes firmly established in the system, it is of course another phoneme like all the rest, and it requires special mention only as regards its history and its relation to the phonemes of neighboring systems. Until it becomes firmly established, however, it is from a descriptive point of view by no means a "phoneme like all the rest." It is in some ways like a free allophone, since [jār] and [jɔ̄r], or [jō̄r] and [jɔ̄r], are in free variation. Yet it is clearly *not* an allophone, since /gār/ and /ō̄r/ show no such variation. With luck, we can even find minimal pairs: '*Mond*' (MHG *māne*) is [mā] or [mɔ̄], [mō̄] or [mɔ̄]; but '*Mann*' (MHG *man*) is always [mā]. We must therefore set up a separate phoneme /ɔ̄/. In historical and dialect studies it is familiar under the name "compromise vowel." Structurally it is characterized by the fact that it *alternates* with some other phoneme, in this case either /ā/ or /ō̄/ or both, and the term "alternating phoneme" would therefore seem appropriate. Another term would do as well; it is only the concept which needs to be recognized.[24]

(6) BALANCE OF AN OPPOSITION. The product of the relative incidences of two phonemes is the balance of the opposition between them. We have seen two examples of the apparent structural effect of imbalance in an opposition. First, the opposition /i ü u/ ≠ /ɪ Ü ʊ/ is consistently maintained in the ZH Highlands, but not consistently maintained in the Lake Area. The reason seems to be that in the Highlands the opposition is in reasonable balance, since the incidence of /i ü u/ includes not only the "special"

[24] Recognition of this concept would go a long way toward resolving the confusion about the status of [i] as a phoneme in American English. It is clear that many speakers have a three-way opposition [ɪ]—[i]—[ʌ], as in [ǰɪst] ('the gist of the matter'), [ǰist] ('just yesterday'), [ǰʌst] ('a just man'). I believe many linguists have hesitated to accept [i] as a phoneme precisely because of the fact that it occurs only as an alternant of (principally) [ɪ] and [ʌ] (though not all examples of [ɪ] and [ʌ] alternate with [i], by any means). Recognizing this special status of [i], and thus not classing it as a "phoneme like all the rest," might settle many a heated argument.

words but also the regular shortenings from MHG /ī ǖ ū/; whereas in the Lake Area the row /i ü u/ is made up almost exclusively of the few "special" words, and hence is quite out of balance with the far higher incidence of /ɪ ʏ ʊ/. Secondly, the northeastern split of MHG /ö o/ into /ö o/≠/ɔ̈ ɔ/ was accompanied by a split of MHG /i ü u/ into /i ü u/≠/e ö o/. It seems likely that this second split occurred because, without it, the opposition /ü u/≠/ö o/ would have been in extreme imbalance, most examples of earlier /ö o/ having been lowered to /ɔ̈ ɔ/.

(7) FUNCTIONAL LOAD. It may seem that the term "balance" of an opposition is unnecessary, since the concept is already covered by the term "functional load" (or "burden," or "yield"). There is, however, a sharp distinction between the two. In the speech of my City of Zürich informant the phonemes /ɛ/ and /e/ are almost in complementary distribution: /ɛ/ occurs almost exclusively before /r x m n ŋ/, and /e/ occurs almost exclusively in other environments. The "functional load" of the opposition /ɛ/≠/e/ is therefore extremely small, since the occurrence of the one or the other phoneme is nearly always predictable in terms of phonological environment. On the other hand, the "balance" of the opposition /ɛ/≠/e/ is reasonably even: the frequency of /e/ (however computed) is certainly the higher of the two, but particularly because of the many lexical items with /ɛr/ the phoneme /ɛ/ is also of common occurrence.

In the case of City of Zürich /i ü u/≠/ɪ ʏ ʊ/, we have an example of small functional load combined with imbalance. As we have seen, the structural result is a confusion of the opposition because, as a contributory factor, the phonetic interval between the two rows is abnormally small.[25] In the case of City of Zurich /ɛ/≠/e/ we have an example of small functional load (because of near complementation) combined with reasonable balance and a normal phonetic interval. The structural result seems to be a strengthening of the opposition. For example, where Weber (*Zürichdt. Gram.*, §308) gives for the older generation /əwǽk/ (beside /əwǣg/) '*weg*' with historically regular /æ/, my younger generation informant says /əwék/ with historically irregular /ɛ/. The incidence of /ɛ/ seems definitely to be on the increase.[26]

[25] One might surmise that the phonetic interval is small because the opposition is out of balance. However, it is clear historically that the very open [ɪ ʏ ʊ] (ranging even to [e ö o]) of the Southwest is an innovation, and that the less open [ɪ ʏ ʊ] of the Center is a preservation of what we know about the earlier phonetic quality of this row.

[26] Such a "strengthening of the opposition" is more familiar when expressed in historical terms. Once the complementary distribution of two allophones has been destroyed in one environment, the two phones—now phonemes—are gradually introduced more and more into other environments as well. For examples, see James W. Marchand, "Germanic short **i* and **e*: Two phonemes or one?", *Language* XXXIII (1957), 346–357, esp. pp. 347–348.

(8) PHONEMIC INDETERMINACY. This concept is hardly new, and yet it is seldom if ever used in phonemic descriptions. Perhaps the reason is that it seems to conflict with the basic tenet of phonemic theory that phonemes are discrete, non-overlapping elements. To admit that one cannot decide whether A and B constitute one phoneme or two would seem to cast reflections on one's abilities as an analyst. Fortunately, dialect study furnishes the theoretical underpinning needed to make phonemic indeterminacy respectable. Phonemic indeterminacy can be admitted if it can be shown that the speech under analysis is at a point of (temporal and/or spatial and/or social) transition from one structure to another. The example cited above in §6 and Map 5 illustrates such a transition: the North and (most of) the East show the single row /i ü u/; the Northwest, Southwest, and Southeast show the opposed rows /i ü u/ ≠ /ɪ ʊ̈ ʊ/; in the Center lies a transitional area of phonemic indeterminacy in which neither of these structures can be fully supported or fully denied.

Princeton University
Princeton, New Jersey

EDWARD STANKIEWICZ

The Consonantal Alternations in the Slavic Declensions

1. According to Sapir, "all languages differ from one another, but certain ones differ far more than others." As for the Slavic languages, he might well have said that they are all alike, but some are more alike than others. The structural similarities stemming from their common origin and from subsequent parallel developments, as well as the divergences between Slavic languages enable us to group them into a number of more or less homogeneous types. Different types may, however, be established, depending on the level of our analysis.[1]

A comparison of contemporary Slavic morphophonemic patterns may reveal that certain types of alternations are common to all Slavic languages, whereas others are alien to them. A productive type of alternation shared by all Slavic languages is, for example, the vowel ~ zero alternation, while consonant ~ zero alternations are rare, and such types as metathesis or reduplication are altogether unknown. Similarities are likewise found in the distribution of certain morphophonemic alternations in different form-classes and grammatical categories. Thus it is apparent that the verb inflection shows different morphophonemic alternations from the nominal inflection. Verbal inflection often admits a variety of concomitant alternations, such as consonantal, vocalic, vowel ~ zero and prosodic, while their co-occurrence is in one way or another restricted in the noun inflection. It should be noted that Slavic verb inflection is formally more complex than the nominal inflection in other ways as well: its desinential suffixes have a more variegated canonical shape than the nominal suffixes, it has a rich system of prefixes which carry grammatical distinctions (of aspect) as well as lexical ones.[2]

[1] On the selection of criteria pertaining to different levels in establishing linguistic types, see the author's "On Discreteness and Continuity in Structural Dialectology," *Word*, XIII (1957), 45.

[2] General problems of morphophonemics are discussed by R. Jakobson in "The phonemic and grammatical aspects of language in their interrelations," *Proceedings of the Sixth International Congress of Linguists*, Paris (1949), pp. 5–18.

Of equal interest are the morphophonemic differences between the separate Slavic languages which are found within the same or similar grammatical categories. For example, the Russian verb inflection admits both accentual and consonantal alternations, but hardly any non-automatic vowel alternations; e.g. /x¹od²it²i/ 'you go' ~ /xad²¹it²i/ 'go' (imperative), /t¹on²it/ '(he) drowns' ~ /tan¹ul/ '(he) drowned', /p²ik¹u/ 'I bake' ~ /p²ič¹oš/ 'you bake', /talkn¹u/ 'I push' ~ /talkn²¹oš/ 'you push', /tku/ 'I weave' ~ /tk²oš/ 'you weave'.[3]

In Polish, on the other hand, where phonemic stress and consequently accentual alternations are excluded, vocalic as well as consonantal alternations are productive in the verb inflection; e.g. /śm²aw/ 'he dared' ~ /śm²eli/ 'they dared', /v²us/ 'he drove' ~ /v²ozwa/ 'she drove' ~ /v²eźli/ 'they drove', /vźow/ 'he took' ~ /vźeli/ 'they took', /vluk/ '(he) dragged' ~ /vlekli/ 'they dragged'. Vowel alternations of analogical origin are in various Polish dialects even more productive than in the standard language; in them we find such pairs as *brał* 'he took' ~ *breli* 'they took', *oblał* '(he) poured' ~ *obleli* 'they poured', *pisał* '(he) wrote' ~ *piseli* 'they wrote'.[4]

Analogical levelling involving vowel alternations has, in the evolution of Polish, moved quite clearly in two opposite directions: vowel alternations became generalized in the verb inflection and underwent various restrictions in the noun inflection. These restrictions account for the formal discrepancies between the modern Polish pairs of nominal forms, such as /b²awi/ (nom. sing.) ~ /b²ali/ (nom. pl.) 'white', /śm²awi/ (nom. sing.) ~ /śm²ali/ (nom. pl.) 'daring', /kšeswo/ (nom. sing.) ~ /kšeśle/ (loc. sing.) 'chair', /žona/ (nom. sing.) ~ /žońe/ (dat.-loc. sing.) 'wife', and their Old Polish equivalents: *biały* ~ *bieli*, *śmiały* ~ *śmieli*, *krzaslo* ~ *krześle*, *żone* ~ *żenie*.

2. The purpose of this paper is to compare the consonantal alternations which occur in the substantive and adjective declensions of the modern

[3] Cf. N. Trubetzkoy, "Das morphonologische System der russischen Sprache," *TCLP*, V: 2 (1934), 29ff. All contemporary Slavic forms are given in this article in phonemic transcription (in slants); alternating forms or phonemes are separated by a tilde; desinential suffixes are transcribed morphophonemically (preceded by a hyphen). The following additional marks have been used: ¹ for stress, / for rising pitch,' for palatal consonants, ² for palatalized consonants,⁻ for length. Dialectal and historical forms are cited in conventional spelling or in transliteration (from Cyrillic). The phonemic interpretation is that accepted by a majority of Slavists; for a survey of the Slavic phonemic systems, see my "Towards a Phonemic Typology of the Slavic Languages," *American Contributions to the Fourth International Congress of Slavicists* (1958), pp. 301–319.

[4] Cf. S. Urbańczyk, *Zarys dialektologii polskiej* (1953), pp. 42–43.

Slavic languages. The following description is primarily concerned with the standard languages; references to sub-standard speech and to dialects is made in cases of morphophonemic divergence from the standard languages. Despite its fragmentary character, the dialectal and sub-standard material is sufficiently illuminating. Neither the scope of this paper nor the present state of Slavic dialectology warrant a fuller presentation of contemporary Slavic morphophonemics.

Even a superficial comparison of the morphophonemic alternations found in the modern Slavic nominal systems with those which are assumed for late Common Slavic reveals that the original system of consonantal and accentual alternations has undergone in almost all Slavic languages deep and diametrically opposite innovations. Only some isolated Slavic dialects (such as Northern Kashubian or some Northwestern Slovenian dialects) exhibit both types of alternations to a similar degree.[5] The morphophonemic innovations have been brought about through phonemic developments of the particular Slavic languages, but even more so through the historical processes known as "analogical levelling."

The discovery of a relationship between the consonantal and accentual alternations in the Slavic nominal systems makes it possible to go beyond a mere enumeration of the morphophonemic rules pertinent to the individual languages. It is possible and methodologically important to show the principles which explain the differences in function and distribution of the morphophonemic alternations in the separate languages. It should be clear that this approach has relevance not only in an interlingual (diachronic or typological) analysis, but in an intra-lingual, synchronic analysis of the sub-systems of individual languages as well. Morphophonemic alternations involve two dimensions: the relationship between various types of alternations within the same language and the relationship between morphophonemic alternations and the grammatical oppositions which they serve to express. They can thus be considered in the two-fold aspect of form and function. In the first case, we ask the question: what is the grammatical load of certain phonemes or distinctive features which alternate with each other, as compared with other alternating phonemes, and second, to what extent are the same phonemes or distinctive features utilized in different form classes or grammatical categories?

[5] The consonantal alternations in Kashubian, which has no literary language, are ignored in this paper. For a description of its accentual patterns see Z. Topolińska's "L'état actuel de l'accent cachoubien," *Slavia*, XXVII:3 (1958), pp. 381–395. For references to consonantal alternations in Slovenian dialects, cf. F. Ramovš, *Morfologija slovenskego jezika* (1952), pp. 44ff, 57ff. An historical survey of alternations in all Slavic languages (from a non-structural standpoint) is found in A. Vaillant's *Grammaire comparée des langues slaves*, 1–3, (1950–58), especially 1, pp. 284–308.

The history of morphophonemic alternations in the Slavic nominal systems and the synchronic facts, which represent the end result of the historical processes peculiar to the separate Slavic languages, make it possible to formulate the answers to these questions both diachronically and synchronically (or typologically). The diachronic formulas which recapitulate the different morphophonemic developments of the Slavic declensions can be phrased as follows: (1) in Slavic nominal systems with accentual (tone and/or stress) alternations, consonantal alternations have been severely restricted or eliminated; conversely, in systems without accentual alternations, the consonantal alternations have remained productive or have undergone further expansion. (2) Reduction of consonantal alternations occurred in the adjectival declensions more often than and prior to that in the substantive declensions, and in the plural of substantives, more often than and prior to that in the singular.

On the synchronic plane, the above formulas can be restated as follows: (1) in Slavic nominal systems with productive accentual alternations, consonantal alternations are absent or unproductive; consonantal alternations are, on the other hand, productive in systems without accentual alternations. (2) Consonantal alternations have more restrictions in the inflection of adjectives than in that of substantives, and more in the plural of substantives than in the singular. The intricate processes of analogical levelling involving consonantal alternations which are listed with meticulous detail in the Slavic historical and comparative grammars are, therefore, to be structurally interpreted as phenomena which led, on the one hand, to the complete or partial liquidation of one set of alternating phonemes (or features) in favor of another set of phonemes, and on the other to the sharpening of the grammatical dichotomies: adjective vs. substantive, plural vs. singular.

3. With respect to the phonemes or distinctive features which participate in the morphophonemic alternations, it is important to emphasize that they cannot be deduced automatically from the phonology of the language in which they occur, but are a matter of selection and morphological assignment. The phonemic system determines only negatively the morphophonemic possibilities of a language: there are no alternations of nonexistent phonemes. However, languages with similar phonemic systems may utilize different morphophonemic alternations. For example, in Russian or in Bulgarian, in which the opposition of hard vs. palatalized consonants is pivotal to the phonemic functioning of these languages, alternations of hard and soft consonants play almost no role in the nominal inflection, although they are of great significance in the verbal

inflection and on the level of derivation. The nominal systems of Standard Serbo-Croatian and of most Serbo-Croatian dialects have grammatically distinctive and productive accentual alternations, but limited consonantal alternations. In some dialects, however (e.g. of Zeta-Lovčen), which have like most Serbo-Croatian dialects phonemic accent and quantity, the accentual alternations in the declension have been impaired, while the consonantal alternations have remained productive.[6] In Czech and Slovak, quantity (which is distinctive in both languages) plays an important morphological role in word formation. In the nominal inflection of Slovak, and even more so of Czech, quantitative alternations are, nevertheless, unproductive. In both languages it is limited to words with monosyllabic stems (which can be listed), and to a small group of "short" adjectives. On the whole, it is possible to say that the feature of length is utilized morphophonemically to a far lesser extent than the prosodic features of tone or of stress. This limited grammatical function of quantity accounts for the fact that the Czech and Slovak declensions are morphophonemically closer to those of the Slavic languages which lack prosodic features in their phonemic systems (and consequently accentual alternations), than to those which have accentual alternations. The above-defined relationship between consonantal and accentual alternations in the Slavic nominal systems is not to be understood as a "law of incompatibility" between the two types of alternation, but rather as a dominant tendency of their historical development and of their present-day structures. While the two types of alternation are mutually exclusive in a number of Slavic nominal systems, others admit both consonantal and accentual alternations. In the latter systems, however, there is a definite hierarchical relationship between the two types of alternations: either the one or the other is productive and largely involved in rendering the nominal distinctions of case, number and gender.

4. If we consider in turn the polarization of the grammatical oppositions adjective vs. substantive and plural vs. singular, we notice that the difference in their morphophonemic complexity (in our case in the admission or exclusion of consonantal alternations) correlates with other characteristics which differentiate them as form classes and as grammatical categories.

In comparison with the substantive declension, the declension of

[6] Cf. P. Ivić, *Die serbokroatischen Dialekte, ihre Struktur und Entwicklung*, 1, (1958), pp. 215ff. Yugoslav dialectologists explain either morphophonemic development as the result of foreign influence. But even if one admits the existence of foreign influence, it is significant that only one type of alternation is eliminated, and that in Serbo-Croatian this is generally the type with consonantal alternations.

adjectives is formally far less differentiated. If we ignore the Slavic dichotomy between "long" and "short" adjectives, which is irrelevant to our problem (and which is, anyway, well preserved only in some Slavic languages), one may say, with Diederichsen, that the adjective which functions as a syntactic modifier or "directed paradigm" is "defective" (i.e. less complex) in comparison with the fundamental or "directing paradigm."[7] The adjectives show greater syncretism of cases than the substantives (e.g. the instrumental-locative singular of non-feminine adjectives in Polish and Upper Lusatian; the genitive-dative-locative singular of feminine adjectives in Polish, Czech and Russian; the genitive-locative plural in all Slavic languages); a reduction of gender distinctions (e.g. in the oblique cases of the plural in all Slavic languages), and less morphophonemic complexity (i.e. the lack or limitation of accentual and/or of consonantal alternations).

In comparison with the singular paradigm, the plural of substantives is likewise formally simplified. In a number of Slavic languages, the plural paradigm shows syncretism of cases which are distinctive in the singular (e.g. the lack of two genitive and locative cases in the plural of Russian substantives; the overlapping of the dative, instrumental and locative in the plural of Serbo-Croatian); it lacks gender or has fewer and different gender distinctions than the singular (e.g. Bulgarian and Russian have no masculine-feminine-neuter distinctions in the plural; Polish and Slovak have in the singular four gender distinctions, and in the plural only the personal-impersonal genders), and it rarely marks gender in the oblique cases. The consonantal alternations which carry or redundantly support grammatical oppositions reflect in most Slavic languages the asymmetry between the singular and plural. The individual languages exhibit in this area some variations; in colloquial Serbo-Croatian, for example, the consonantal alternations are more stable in the plural than in the singular, while the accentual alternations are more variegated and stable in the singular.

Morphophonemic alternations which represent one of the formal expressions of the grammatical complexity of a language, can thus not be separated from its overall grammatical structure; they reflect the morphological relationships and are themselves carriers of these relationships.

5. In considering the functions of the consonantal alternations, it is important, furthermore, to examine them with relation to the desinential suffixes. While the consonantal alternations involve only some types of

[7] Cf. P. Diderichsen, *TCLC, Recherches Structurales*, V (1949), 145; J. Kuryłowicz, *L'accentuation des langues Indo-Européens*, 2nd ed. (1958), pp. 215–216.

stems (terminating with particular consonants) and certain cases of the declension, the desinential suffixes constitute the obligatory, primary carriers of grammatical oppositions. In cases in which the suffixes have distinctive phonemic forms, the alternations are concomitant, and may also signalize morpheme boundaries. In those cases, however, which have homonymous suffixes, the alternations themselves carry grammatical oppositions. I will call the former alternations redundant and the latter distinctive. Distinctive alternations are, for example, the alternations of non-rising ~ rising pitch in the dative vs. locative singular in Serbo-Croatian (e.g. /grādu/ vs. /grȃdu/ 'town', /nosu/ vs. /nósu/ 'nose'), the alternations of stress in the genitive singular vs. nominative (-accusative) plural in Russian (e.g. /izbʲi/ vs. /ʲizbi/ 'cottage', /pˀisˀmʲa/ vs. /pˀʲisˀma/ 'letter', /glʲaza/ vs. /glazʲa/ 'eye') or the consonantal alternations in the nominative plural of Czech or Polish substantives, where the shape of the final stem-consonant may determine gender differences (e.g. Czech /kmotři/ 'godfathers' vs. /kmotri/ 'godmothers'; Polish /swuʒi/ 'male servants' vs. /swugˀi/ 'maid servants', /češi/ 'Czechs' vs. /čexi/ 'Bohemia'). Examples of redundant alternations are found in the nominative and dative-instrumental-locative plural of Serbo-Croatian masculine substantives, which in contradistinction to the respective forms of feminine substantives have also different desinential suffixes (e.g. /drūzi, drūzima/ [or /drugovi, drugovima/] 'male companions' vs. /drūge, drūgama/ 'female companions'). The redundant alternations are far from irrelevant in the functioning of morphological systems, for they often serve as features of differentiation between generic grammatical categories.[8] Russian stress alternations, for example, are in most cases of the declension, redundant, but their basic function is to sharpen the polarity between the singular and the plural (e.g. /boj, bʲoja, bʲoju/ vs. /bajʲi, bajʲof, bajʲam/ 'battle'; /dušʲa. dušʲi, dušʲe/ vs. /dʲuši, duš, dʲušam/ [or /dušʲam/] 'soul'; /slʲova, slʲovu/ vs. /slavʲa, slof, slavʲam/ 'word'). In the history of the Russian declension there were many kinds of accentual levelling which led to the sharpening of this opposition.[9]

Some morphophonemic alternations have, however, no such clear morphological function, but are distributed, as if haphazardly, in specific

[8] The terms "generic" and "specific category" are defined by Hockett as follows: "a generic category is a whole system of classification"; "a specific grammatical category is an element in a system or a class in a classification." Cf. C. F. Hockett, *A Course in Modern Linguistics* (1958), p. 231.

[9] For a history of accentual levelling in the Russian declension, cf. L. Tesnière, "L'opposition morphologique de l'accent dans le substantif russe," *Mélanges en l'honneur de Jules Legras* (1937), p. 1ff.

grammatical categories. (In Trubetzkoy's terminology, these are "irrational" alternations.)[10] Such alternations frequently result when automatic, phonemically conditioned alternations change into non-automatic, morphologically conditioned alternations as a result of phonemic change. An example are the C ~ C' alternations in the Eastern Ukrainian dialects (and in Eastern Standard Ukrainian), which arose after the coalescence of the phonemes /i/ and /y/. These alternations are of recent date and quite complicated.[11]

Besides distinctive or redundant functions, the presence or absence of morphophonemic alternations may also serve to render expressive (emotive) differences. In Polish, the presence of a consonantal alternation in the nominative plural of personal substantives is emotionally neutral, whereas the lack of an alternation in the same form carries a pejorative connotation (e.g. /doktoži/ 'doctors', /advokaći/ 'lawyers', /pʾijaci/ 'drunkards', /bwaźńi/ 'court jesters', as opposed to /doktori/, /advokati/, /pʾijakʾi/, /bwazni/; the latter may be roughly translated as 'horse doctors', 'shysters', 'lushes', 'clowns').[12]

It will be observed that in those languages in which the consonantal alternations are productive, they are frequently also endowed with distinctive functions. For example, in Polish they may render the oppositions of gender, number and case; e.g. /markʾiźi/ (nom. pl. personal) 'marquises' vs. /markʾizi/ (nom. pl. impersonal) 'marchionesses';[13] /dobri/ (nom. sing. masculine) vs. /dobži/ (nom. pl. personal); /žońe/ (dat. sing. fem.) vs. /žone/ (acc. sing. fem.) 'wife'. On the other hand, in the languages in which the consonantal alternations are restricted, their functions are redundant in most cases; e.g. Bulgarian /vəlk/ (sing. masc.) bs. /vˡəlci/

[10] Op. cit., pp. 34–35.

[11] The C ~ C' alternations in Eastern Ukrainian involve final stem consonants which were historically followed by /y/ and are hard vs. final stem consonants followed by /i/ which are presently soft; thus the soft consonant occurs in the loc. sing. and in the gen. plur. (before the suffix *-iv*) of II declension substantives, and the hard consonant elsewhere; e.g. /jˈavorʾi, jˈavorʾiv/ vs. nom. plur. /jˈavori/, dat. plur. /jˈavoram/; in the III declension the hard consonant occurs in the gen., vocative sing. and in the gen. plur. and the soft consonant elsewhere; e.g. gen. sing. /pˈovʾisti/, gen. plur. /pˈovʾistej/ vs. nom. sing. /pˈovʾistʾ/, nom. plur. /pˈovʾistʾi/; in the I and "mixed" declension, the soft consonant occurs only in the dat.-loc. sing. and the hard consonant elsewhere; e.g. dat.-loc. sing. /orlʾˡatʾi/ vs. gen. sing. /orlʾˡati/, nom. plur. /orlʾˡata/.

[12] A more detailed treatment of this problem in Polish is given in my "Distribution of Morphemic Variants in the Declension of Polish Substantives," *Slavic Word* IX (1955), 554ff.

[13] It must be remembered that gender is generally inherent in the stem of substantives, and that there are few examples in which gender is differentiated only through morphophonemic alternations.

(pl.) 'wolf', or Serbo-Croatian /drūzi/ (nom. pl. masc.) 'male companions' vs. /drūge/ (nom. pl. fem.) 'female companions'. In both examples the suffixes themselves would suffice to distinguish number and gender. In these languages the consonantal alternations may, however, set apart the rare cases when adjectives and substantives have homonymous stems; e.g. Bulgarian /samuˈluki/ (plur. adj.) vs. /samuˈluci/ (plur. substantive) 'self-taught'. It is, further, noteworthy, that while the lack of consonantal alternations may function in a language like Polish as an expressive device, it is, on the contrary, the presence of alternation which fulfills a similar function in languages with limited alternations; e.g. Bulgarian/ cˈigani/ ("neutral" plur.) vs. /ciganˀˈa/ (pejorative plur.), from /cˈiganin/ "Gypsy."

6. The assumption that language is a system and that its parts are interrelated and explainable in the light of the whole, is one of the cornerstones of modern linguistics, even if the notion of interdependence is given a different interpretation by various linguists. It is impossible to grasp the functioning of morphophonemic alternations apart from concrete systems. Nor is it possible to explain, on the diachronic plane, the direction of the processes known as "analogical levelling." The atomistic approach to the problem of grammatical levelling has no doubt been the main reason for the failure of so many attempts to explain their role in the history of grammatical systems. The Neogrammarians have generally adopted an "ignorabimus" attitude as to the direction and causes of analogical levelling. Not only were they and their followers lost in recording the minutiae of analogical change, but they programmatically rejected theoretical concepts in favor of the empirical task of fact gathering. The Gospel of the Neogrammarians, Paul's *Prinzipien der Sprachgeschichte*, declares that "to show why analogy works in one language in one way and in another in another way, we would have to be omniscient."[14] Within the framework of their psychological approach to language, analogy was also interpreted as "accidentale Vorgänge" in the idiolects of individuals. Some Neogrammarians, like Wundt, operated with the concept of mechanistic sound associations ("associative Fernwirkung der Laute"), or with the statistical notion that forms with a higher frequency of occurrence eliminate ("induce") forms with a lower frequency of occurrence. Opponents of the statistical approach (especially Brugmann, Kruszewski and Baudouin de Courtenay) have shown that the opposite phenomenon is equally common. Some linguists (including de Saussure and some of his followers) tend to explain analogical levelling as a corrective to the "blindness of phonetic

[14] H. Paul, *Prinzipien der Sprachgeschichte*, 1st ed. (1880), p. 208.

laws." Phonology and morphology are viewed (diachronically) not as two independent, though interacting, levels, but as two conflicting orders, one of which mends what the other destroys. The effect of analogical levelling is thus viewed as a re-establishment ("Herstellung," in Paul's terms) of the original grammatical order. As has been pointed out by Žirmunskij,[15] the role of grammatical levelling is frequently discussed in such vague and abstract phraseology as "tendency towards symmetry," "unification," "harmony," "Systemzwang," without any indication of how it operates in concrete systems and how it correlates with other formal elements in a language. Žirmunskij himself sees in the work of analogy a sort of teleological striving towards progress. "The progressive character of analogy," he says, "is to improve the grammatical rules of a language and to perfect the functioning of its grammatical system."[16] It is hard to say to what extent Žirmunskij pays lip service to official doctrine. But he himself would probably hesitate to call all change progress. And grammatical levelling, like other linguistic change, may either simplify or complicate, or merely transform one type of linguistic economy into another. Since grammatical levelling at the same time participates in and carries the morphological hierarchy of a language, it is the task of the linguist to indicate the functioning of co-existing elements within the total system, their change in function, or complete elimination. Aside from the theoretical difficulties involved in teleological explanations, these always seem to be a posteriori statements when applied to linguistic diachrony.[17]

An attempt to establish general laws of analogical change has been made in recent years by Kuryłowicz.[18] These have been subjected to a thorough critical analysis by Mańczak,[19] whose copious examples clearly contradict these "laws." Mańczak himself, however, returns to the very statistical

[15] V. M. Žirmunskij, "Vnutrennie zakony razvitija jazyka i problema grammatičeskoj analogii," *Trudy Instituta Jazykoznanija ANSSR*, 4, (1954), pp. 74ff. Žirmunskij's article contains also an exhaustive survey of the history of the problem and a rich bibliography. For complementary bibliography and some comments on Žirmunskij's article, see L. A. Bulaxovskij, "Grammatičeskaja indukcija v slavjanskom sklonenii," *Voprosy Jazykoznanija*, (1956), no. 4, 14ff., and idem, "Issledovanija v oblasti grammatičeskoj analogii i rodstvennyx javlenii," *Učenye Zapiski Xar'kovskogo Universiteta*, XIX (1940), 1–31.

[16] Op. cit., p. 110.

[17] On the difficulties of teleological explanations, cf. I. Scheffler, "Thoughts on Teleology," *The British Journal for the Philosophy of Science*, IX, 36 (1959), pp. 265–283; and in linguistics, E. Coseriu, *Sincronía, Diacronía e Historia* (1958), pp. 101ff.

[18] J. Kuryłowicz, "La nature des procès dits analogiques," *Acta Linguistica* V (1951), 15ff.

[19] W. Mańczak, "Tendances générales des changements analogiques," *Lingua*, VII (1958), 298ff.; 403ff.

position which Kuryłowicz correctly rejected, and his "Hypotheses" do not survive critical examination. To prove, for example, that alternations within the stem are more frequently eliminated than introduced, Mańczak cites the fate of the vowel alternations in the history of Polish substantives.[20] But a survey of the vowel alternations in the evolution of the Polish verb would have led him to the opposite conclusion. It is obvious, therefore, that statistical or universal laws of analogical levelling are either so general as to be trivial, or do not stand the test of empirical validation.

7. The consonantal alternations which occur in the Slavic substantive and adjective declensions fall into three types which can be expressed in the formulas (1) K ~ C, (2) T ~ Ḱ, and (3) C ~ C°. Alternations of velars with palatals (which resulted historically from the so-called "first palatalization") are rare in the Slavic declensions.[21] Type 1 involves the alternation of velar consonants (plus *h*, or *h* instead of *g* in some languages) with consonants, which resulted from the so-called "second palatalization," the reflexes of which are somewhat different in the separate Slavic languages. In the Eastern and Southern Slavic languages (and in Slovak), *k*, *g*, (*h*), *x* alternate with *c*, *z*, *s*; in the Western Slavic languages, *k*, *g*, (*h*), *x* alternate with *c*, *ʒ* (or *z*), *ś*; in Polish, *x* alternates with *ś* in the nominative plural. Type 2 involves basically alternations of dental and palatal consonants which are productive only in the Western Slavic languages (cf. §9). Type 3 involves alternations of hard and palatalized consonants.

The three types of alternations under consideration are distributed in the following cases of the modern Slavic declensions, with some variations peculiar to the individual languages:

(a) in the dat.-loc. sing. of substantives of the I (predominantly feminine) declension, and to a lesser extent in the loc. sing. of the II (non-feminine) declension.

(b) in the nom. plur. of masculine substantives (or, simply, in the plural of those languages which have no case systems).

(c) in the nom. plur. (or simply plur.) of adjectives.

(d) in the dat.-instr.-loc. plur. of II declension substantives in Serbo-Croatian; in the loc. plur. of substantives in Czech and sporadically in Polish.

8. The languages with tone or stress alternations, in which the consonantal alternations have been variously impaired, include the Eastern

[20] Op. cit., pp. 301ff.

[21] In most Slavic languages, the alternation /k, x/ ~ /č, š/ concerns residual forms: e.g. Ukr. /'oko, 'uxo/ 'eye, ear' ~ /'oči, 'uši/, Russ. and BR. /klok/ 'tuft' ~ /kl'očja/, resp. /kl'očča/.

Slavic languages (Russian, Ukrainian, Byelorussian), and the Southern Slavic languages (Slovenian, Serbo-Croatian and Bulgarian), with the exception of Standard Macedonian, which lacks prosodic features (and consequently accentual alternations).

The consonantal alternations occurring in the nominal inflections of these languages will be discussed in order of their increasing productivity. The following types of alternation may occur: (1) the K ~ C alternation, (2) the C ~ C^{j} alternation, which is relevant only for the Eastern Slavic languages and for Bulgarian, which have this phonemic opposition.

Neither of the two types is productive in the modern Russian declension. In Standard Russian there are only a few words with a C ~ C^{j} alternation; i.e. /čort/ 'devil', /mud^{l}o/ 'testicle', /sas^{j}^{l}et/ 'neighbor', /kal^{j}^{l}ena/ 'knee' vs. nom. plur. /č^{l}ert^{j}i, m^{l}ud^{j}i (or /mud^{j}^{l}e/), sas^{j}^{l}ed^{j}i, kal^{j}^{l}en^{j}i/. (The first two are taboo words.) The K ~ C alternation, which disappeared around the fifteenth century (and even earlier in Northern Russian), is found residually only in *druk* 'friend' vs. nom. plur. /druz^{j}j^{l}a/. In eighteenth and nineteenth century Standard Russian, there were still a few short adjectives with a C ~ C^{j} alternation in the nom. plur. (e.g. *r^{l}ad^{j}i* 'glad', *gar^{l}azd^{j}i* 'clever') which have subsequently lost the alternation.[22]

In standard Slovenian there are only two substantives with a K ~ C alternation; i.e. /volk/ 'wolf', /otrok/ 'child' vs. nom. plur. /volcj^{l}e, otrôci/. K ~ C alternations are preserved only in some peripheral Northwestern dialects.[23]

In Standard Serbo-Croatian, which is based on the Hercegovina dialect, the K ~ C alternations are subject to various restrictions. In the dat.-loc.-sing. of I declension substantives, the alternation is regular only for *k*, *g* ~ *c*, *z*; e.g. /rȗka/ 'hand', /nóga/ 'leg' ~ /rūci, nozi/, whereas a final x does not regularly alternate; e.g. /snáxa/ 'daughter-in-law', /muxa/ 'fly' ~ /snáxi/ (or /snái/, rarely /snási/), /muxi/ (or /mui/). In case of consonantal clusters, the alternation is exceptional; e.g. /kocka/ 'die', /mázga/ 'mule', /bitka/ 'battle', /mačka/ 'cat', /guska/ 'goose' ~ /kocki, mázgi, bitki, mački, guski/. Hypocoristica, loanwords and proper names likewise exclude alternations; e.g. /séka/ 'sister', /bȁka/ 'grandmother', /āga/ 'agha', and proper names /raka, volga/ ~ /séki, bȁki; āgi; raki, volgi/. In the colloquial language, this alternation is eliminated; thus we find

[22] On the history of the consonantal alternations in Russian, see A. V. Isačenko, "Der grammatische Wechsel k/c, g/z im Russischen," *Slavia*, XIV (1935–36), 43–44; L. A. Bulaxovskij, *Istoričeskij kommentarij k russkomu literaturnomu jazyku*, 5 ed. (1958), pp. 189ff.; P. Ja. Černyx, *Istoričeskaja grammatika russkogo jazyka* (1954), pp. 139, 174–75.

[23] Cf. footnote 5.

/réka, slúga, snáxa/ ~ réki, slúgi, snáxi (or snái)/. The K ~ C alternation occurs furthermore in the nom. plur. of masculine substantives (before the ending *-i*) and in the dat.-instr.-loc. plur. of masculine and of three (bookish) neuter substantives (before the ending *-ima*); e.g. /júnāk/ 'hero', /bíl'eg/ 'mark', /sirómax/ 'pauper'; /blāgo/ 'treasure' ~ (nom. plur.) /junáci, bíl'ezi, sirómasi/; (dat.-instr.-loc. plur.) /junácima, bíl'ezima, sirómasima/; /blāzima/. Since the *-i*, *-ima* suffixes are rarely attached to monosyllabic or to certain types of polysyllabic stems (which take the suffixes *-ovi*, *-ovima*), the K ~ C alternation is limited also in the plural of masculine substantives. Within the substantive declension, the alternation is only redundant. The declension of Serbo-Croatian adjectives excludes consonantal alternations. In the Čakavian, Kajkavian and in most Štokavian dialects consonantal alternations were altogether eliminated before the sixteenth century.[24]

In Bulgarian, which has no case system, the alternations concern only the plural and collective forms of substantives. Adjectives admit no alternations. What seem to be C ~ C^{j} alternations in the plural of some substantives are cases of automatic alternation (i.e. of the phonemic neutralization of soft consonants before front vowels and in final position), as in (plur.) /z^{l}et^{j}uvi, ugn^{j}^{l}ovi, mur^{j}^{l}a/ vs. (sing.) /zet/ 'son-in-law', /^{l}ogən/ 'fire', /mur^{l}e/ 'sea'. The non-automatic C ~ C^{j} and K ~ Č alternation is found in a small number of collective forms, as in /n^{l}iva/ 'field', /k^{l}əšta/ 'house' ~ /n^{l}iv^{j}a, kəšt^{j}^{l}a/, and in some "pejorative" plural forms, as in /grək/ 'Greek' ~ /gərč^{l}i/ or /gərčol^{j}^{l}a/, /'b^{l}ivul/ 'buffalo' ~ /bivul^{j}^{l}a/, /c^{l}iganin/ 'Gypsy' ~ /cigan^{j}^{l}a/ (the corresponding "neutral" plurals are /g^{l}ərci, b^{l}ivuli, c^{l}igani/). The K ~ C alternation is rare in masculine monosyllabic substantives, since most of them attach the suffix *-ove*, which excludes alternation in the stem, and not *-i*, which admits it; some of the forms which take the latter suffix likewise admit the suffix *-ove* as a free variant; e.g. /vnuk/ 'grandchild', /vəlk/ 'wolf' ~ /vn^{l}uci, v^{l}əlci/; /bik/ 'bull', /zvuk/ 'sound', /vrak/ 'enemy' ~ /b^{l}ici/ or /b^{l}ikuvi/, /zv^{l}uci/ or /zvukuv^{l}e/, /vraz^{l}i/ or /vraguv^{l}e/. The K ~ C alternation is also found in two feminine nouns: /rək^{l}a/ 'hand' and archaic /nug^{l}a/ 'leg' ~ (plur.) /rəc^{l}e, nuz^{l}e/.

In the polysyllabic masculine stems, the K ~ C alternation shows considerable oscillation, with a tendency to disappear in the colloquial

[24] Cf. H. Rubenstein, *A Comparative Study of Morphophonemic Alternations in Standard Serbo-Croatian, Czech and Russian* (1950); A. Meillet and A. Vaillant, *Grammaire de la langue Selbo – Croate*, 6 ed. (1952), pp. 86 ff., A. Leskien, *Grammatik der serbo-kroatischen Sprache* (1914), p. 344; P. Ivić (op. cit.), esp. pp. 143, 192, 235, 259, 277, 297.

language. But even in the standard language, where the alternation owes its durability greatly to modern normative attitudes (it was less common in the works of 19th c. Bulgarian writers), it has various restrictions in its distribution; thus it cannot occur after *x*, *sk*, *ng* and in loanwards; e.g. /orˈex/ 'nut', /ˈotpusk/ 'leave', /mˈitink/ 'rally', /ˈdˈansink/ 'dancing', /zigzˈak/ 'zigzag' ~ (plur.) /orˈexi, ˈotpuski, mˈitingi, dˈansingi, zigzˈagi/. In the colloquial language, we find such plural forms as /trˈaki/ 'Thracians', /rˈaki/ 'crabs', /frˈanki/ 'francs', /mustˈaki/ 'mustaches' (sing. /trak, rak, frank, mustˈak/). The lack of alternations has also been recorded in various Bulgarian dialects.[25]

In the nominal systems of standard Ukrainian and of Byelorussian, the consonantal alternations are likewise limited, but have a somewhat different range of distribution from that in the above languages. The C ~ C' alternations are automatic in Western Ukrainian, but not in Eastern Ukrainian or in Byelorussian. A non-automatic C ~ C' alternation involving the phonemes /t, d/ ~ /c', ʒ'/ is found in Byelorussian in /čort/ 'devil', /susˈet/ 'neighbor' ~ (plur.) /čˈerc'i, susˈeʒ'i/ (cf. the Russian /čˈert'i, sasˈed'i/).

The K ~ C alternation does not occur in the plural of substantives, nor in the adjectival declension of either Ukrainian nor Byelorussian. In both languages it is productive only in the dat.-loc. of first-declension substantives where it is not distinctive; e.g. Ukr. /rukˈa/ 'hand', /nohˈa/ 'leg', /mˈuxa/ 'fly' ~ (dat.-loc.) /rucˈi, noz'ˈi, mˈus'i/; BR. /rukˈa/ 'hand', /nahˈa/ 'leg', /straxˈa/ 'thatch' ~ (dat.-loc.) /ruc'ˈe, naz'ˈe, stras'ˈe/. In the loc. sing. of second-declension substantives, there is considerable oscillation. In Ukrainian the alternation is restricted to simple masculine substantives and to a few neuter substantives; e.g. /mox/ 'moss', /v'ik/ 'age', /kˈelex/ 'chalice'; /ˈoko/ 'eye'. /vˈuxo/ 'ear', /jˈabluko/ 'apple' ~ (loc.) /v'ˈici, mˈos'i, kˈeles'i; ˈoci, vˈus'i, jˈabluci/. Most masculine substantives however, (especially those with a movable accent) admit variants without the consonant alternation; e.g. /verx/ 'top', /strax/ 'fear', /torh/ 'market', /b'ik/ 'side' ~ (loc.) /vˈers'i, strˈas'i, tˈorz'i, bˈoci/ (or /verxˈu, straxˈu, torhˈu, bokˈu/). That the alternation is unproductive is shown by the fact that it is absent from derivative forms and from loanwords; e.g. /l'int'ˈux/ 'lazybones', /kožušˈok/ 'fur', /smˈoking/ 'smoking', /mˈolox/ 'moloch', /morh/ 'morgue' ~ (loc.) /l'int'ˈuxu, kožuškˈu, smˈokingu, mˈoloxu, mˈorhu/. A similar situation exists in Byelorussian, with the difference that the second-declension substantives with alternating con-

[25] R. Bernard, "Quelques observations sur la seconde palatalization en bulgare moderne," *Sbornik v čest na Aleksand'r Teodorov-Balan* (1956), pp. 85ff.; L. Beaulieux, *Grammaire de la langue bulgare* (1950), pp. 33ff.

sonants are even less numerous (they occur mainly in stems with a final *h*); e.g. /box, sn^{o}ex, rox; br^{l}uxa, j^{l}ablika/ 'God, snow, horn; belly, apple' ~ (loc.) /b^{l}oz^{o}i, sn^{o}^{l}ez^{o}i/ (or /sn^{o}ah^{l}u/), /r^{l}oz^{o}i/; /br^{l}us^{o}i, j^{l}ablici/ or /br^{l}uxu, j^{l}abliku/. The *-u* ending and the lack of alternation in the loc. sing. of II declension substantives is general in various Byelorussian dialects.[26] The nominative plural of substantives has a K ~ C alternation only in one Ukrainian dialect. This is the Lemki dialect, which is the only one in Ukrainian without phonemic stress (and consequently without accentual alternations).[27]

In the Eastern Ukrainian dialects, where the loss of the phoneme /y/ is responsible for a new complex pattern of C ~ C^{o} alternations in the declension of substantives, it is interesting to notice the phenomenon of free variation in the final consonants of adjectival stems; the tendency is, however, to generalize either the hard or the soft consonant throughout the declension of adjectives. Thus we find dialects which admit only final hard consonants (e.g. /v^{o}^{l}irni/ 'faithful', /b^{l}osi/ 'barefoot', /mal^{l}i/ 'small', /molod^{l}i/ 'young' in the nom. sing. and nom. plur. of "contracted" adjectives), or only final soft consonants (e.g. /v^{o}^{l}rn^{o}i, b^{l}os^{o}i, mal^{o}^{l}i, molod^{o}^{l}i/).[28]

9. Accentual alternations are absent from the Western Slavic languages and from Macedonian. Most of these languages have productive consonantal alternations, which include, besides the K ~ C and C ~ C^{o} alternations, also the T ~ Ḱ alternation. C ~ C^{o} alternations are absent from Czech, Slovak, and Macedonian, which lack the feature of palatalization. The T ~ Ḱ formula involves in the particular languages the following alternations of consonants: 1. in Polish: *t, d, n, s, z* ~ *ć, ʒ́, ń, ś, ź*; *w, r* ~ *l, ž* and *š, ž* ~ *ś, ź* (the last only in the nom. plur. of adjectives); 2. in Czech and Upper Lusatian: *t, d, n* ~ *t'* (=U.Lus. *ć*), *d'* (=U.Lus. *ʒ́*), *ń*; *r* ~ *ř*; in Upper Lusatian also *w* ~ *l*. 3. in Slovak: *t, d, n* ~ *t'*, *d'*, *ń* and *l* ~ *l'* (the phonemic opposition *l*/*l'* is optional). 4. in Lower Lusatian: *t, d, n* ~ *ś, ź, ń*; *r* ~ *r^{o}*; *w* ~ *l*.

The T ~ Ḱ alternation in these languages involves not only the final

[26] The oscillations in the Ukrainian consonantal alternations are briefly stated in J. Šerex's *Narys sučasnoji Ukrajins^{o}koji literaturnoji movy* (1951), pp. 191–192; for Byelorussian, cf. T. P. Lomtev, *Grammatika belorusskogo jezyka* (1956), pp. 67ff.; and for the history and dialects, M. A. Žydovič, "Da historyi sklanennja nazoŭnikaŭ u belaruskaj move," *Pracy Instituta Movaznaŭstva ANBSSR*, Vypusk III(1957), esp. pp. 3–31. In the case of Ukrainian and Byelorussian, the author profited greatly from native informants.

[27] Cf. F. T. Žylko, *Narysy z dialektolohiji ukrajins'koji movy* (1955), p. 153.

[28] F. T. Žylko (op. cit.), esp. pp. 165, 174.

consonant of the stem, but also certain consonantal clusters (e.g. *sk, ck* ~ *št', čt'*, in Czech; *st, zd, sn, zn* ~ *ść, źʒ, śń, źń* in Polish). The numerous and variegated consonantal alternations are in all Western Slavic languages endowed with distinctive and redundant functions which express the oppositions of gender, number and case.

The consonantal alternations in the nominal inflections of these languages are discussed in order of their decreasing productivity.

In the Polish singular declension, the K ~ C alternation occurs in the dat.-loc. sing. of I declension substantives; the other types of alternation involve both the I and II declension substantives (the latter only in the loc. sing.); e.g. /fizika/ 'physics', /toga/ 'toga', /sčšexa/ 'thatch' ~ (dat.-loc.) /fizice, toʒe, sčšeše/; /mˀut/ 'honey', /moda/ 'fashion', /stado/ 'herd'; /buta/ 'arrogance', /śfˀento/ 'holiday'; /stuw/ 'table', /kowo/ 'wheel'; /kora/ 'bark', /profesor/ 'professor' ~ (dat.)-loc. /mˀoʒe; moʒe, staʒe; buće, śfˀeńće; stole, kole, kože, profesoże/; /atom/ 'atom', /mˀikrop/ 'microbe', /bˀiskup/ 'bishop', /garderoba/ 'wardrobe', /swoma/ 'straw' ~ (dat.)-loc. /atomˀe, mˀikrobˀe, bˀiskupˀe, garderobˀe, swomˀe/. In addition to the above alternations, Polish has also a *k* ~ kˀ and *g* ~ gˀ alternation in the accusative singular of I declension substantives, which until recent times was automatic.[29] Consonantal alternations are unproductive in the loc. plur. of substantives; they are found only in a few geographical names; e.g. /vengri/ 'Hungary', /prusi/ 'Prussia', /vwoxi/ 'Italy' ~ (loc.) /vengžex, prušex, vwošex/. All three types of alternation occur in the nom. plur. of personal substantives; e.g. /fizik/ 'physicist', /kolega/ 'colleague', /mńix/ 'monk' ~ (nom. plur.) /fizici, koleʒi, mńiši/; /idjota/ 'idiot', /sǫšat/ 'neighbor', /rabˀin/ 'rabbi', /doktur/ 'doctor', /ańow/ 'angel' ~ (nom. plur.) /idjoći, sǫšeʒi, rabˀinˀi, doktoži, ańeli/; /bˀiskup/ 'bishop', /snop/ 'snob', /agronom/ 'agronom' ~ (nom. plur.) /bˀiskupˀi, snobˀi, agronomˀi/. Although the suffix *-i* (with preceding consonantal alternation of the stem) is, in some environments, mutually exclusive or in free variation with the suffix *-ovˀe* (without stem alternation), it is obligatory in other environments (i.e. after dental consonants and *k*) and in some lexical items; the productivity of the consonantal alternations in the nom. plur. is thus insured. All three types of alternation occur also in the nom. plur. of personal gender adjectives, which admit in addition an alternation of the palatal spirants *š, ž* ~ *ś, ź* (strident vs. mellow), which is of purely analogical origin; e.g. /visokˀi/ 'high', /drogˀi/ 'dear', /śfˀenti/ 'holy', /tfardi/

[29] For a more detailed statement of the Polish consonantal alternations, see the author's "The Distribution . . ." (op. cit.). The history of the Polish declension and of its alternations are outlined in *Gramatyka historyczna języka polskiego*, by Z. Klemensiewicz, T. Lehr-Spławinski, S. Urbańczyk (1955), pp. 266ff.

'hard', /dobri/ 'good', /koxani/ 'beloved', /wisi/ 'bald', /pʾeši/ 'pedestrian', /duži/ 'big' ~ (nom. plur.) /visoci, drozi, śfʾeńći, tfarʒi, dobži, koxańʾi, wiśi, pʾeśi, duźi/.

Equally productive are the consonantal alternations in Upper and Lower Lusatian. They occur in the dat.-loc. sing. and in the nom.-acc. dual of first-declension substantives; e.g. U.Lus. /ruka/ 'hand', /noha/ 'leg', /tšẽxa/ 'thatch'; /žona/ 'wife', /pata/ 'heel', /voda/ 'water'; /swoma/ 'straw', /kupa/ 'island', /kruva/ 'cow' ~ (dat.-loc. sing. and dual) /ruce, noze, tšẽše; žońe, paće, voʒe; swomʾe, kupʾe, kruvʾe/. Low. Lus.: /ruka/ 'hand', /droga/ 'road', /muxa/ 'fly'; /śota/ 'aunt', /gvʾẽzda/ 'star', /voda/ 'water', /živa/ 'vein'; /riba/ 'fish', /zima/ 'winter' ~ (dat.-loc. and dual) /ruce, droze, muše; śośe, gvẽźźe/ (spelled *gwězdźe*), /voźe, žile; ribʾe, zimʾe/.

The C ~ Cʾ and T ~ Ḱ alternations are in both Lusatian languages the rule in the loc. sing. of II declension substantives; e.g. U.Lus. /hrôt/ 'castle', /lut/ 'people', /ʒẽwo/ 'work', /zwoto/ 'fold'; /zup/ 'tooth' ~ (loc.) /hrôʒe, luʒe, ʒẽle, zwoće; zubʾe/; Low.Lus. /svʾẽt/ 'world', /mʾẽsto/ 'city'; /dup/ 'oak' ~ (loc.) /svʾẽśe, mʾẽśśe (spelled *měsće*), /dubʾe/. There are also masculine and some neuter substantives which have a K ~ C alternation in the loc. sing., although this alternation is no longer compulsory; e.g. U.Lus. /śńẽk/ 'snow', /mloko/ 'milk' ~ (loc.) /śńẽze, mloce/; Low.Lus. /jabwuko/ 'apple', /śńẽk/ 'snow', /prox/ 'powder' ~ (loc.) /jabwuce, śńẽze, prose/. Consonantal alternations are the rule in the nom. plur. and collective of substantives and of adjectives of the personal gender. Thus the nom. plur. personal forms of the U.Lus. substantives /vojak/ 'soldier', /čex/ 'Czech', /pohan/ 'pagan', /posow/ 'delegate', /socjalist/ 'socialist', /student/ 'student', are /vojaci, češi, pohańi, pôsli, socjalisći, studenći/, while the collective forms of /bratr/ 'brother', /mńix/ 'monk', /agent/ 'agent', /žit/ 'Jew' are /bratřa, mńiša, agenća, žiʒa/; and the nom. plur. of the Low. Lus. substantives /kamarat/ 'comrade', /suset/ 'neighbor', /cart/ 'devil', /bratš/ 'brother', /cigan/ 'Gypsy' are /kamaraśi, suseźi, carśi, bratśi, cigańi/. Examples of personal adjectives in the nom. plur. are in U.Lus.: /vuwci/ 'big'. /drozi/ 'dear', /hwusi/ 'deaf', /bohaći/ 'rich', /horʒi/ 'proud' (nom. sing. /vuwki, drohi, hwuxi, bohati, hordi/), and in Low. Lus.: /daloci/ 'remote', /drozi/ 'dear', /tvarźi/ 'hard '(nom. sing. /daloki, drogi, tvardi/). The *-i* suffix of personal substantives (with stem alternation) is in Upper Lusatian in complementary distribution or in free variation (in some environments only) with the personal suffix *-ojo/-o*.[30]

[30] Cf. P. Wowčerk, *Kurzgefasste obersorbische Grammatik* (1954); B. Šwela, *Grammatik der niedersorbischen Sprache* (1952). Neither grammar presents an adequate picture of the consonantal alternations in Lusatian.

The consonantal alternations in Czech are somewhat more complex than in Polish or Lusatian, because of the greater variations encountered in the standard language and in the Czech dialects. The situation is further complicated by the fact that Czech is sharply divided into two social dialects: i.e. literary Czech (*spisovná čeština*) and colloquial Czech (*obecná čeština*). The morphophonemic patterns of these two dialects show, like their phonemic systems, serious divergences.

The K ~ C and T ~ Ḱ alternations are found in the following environments: (a) in the dat.-loc. sing. of first-declension substantives; e.g. /ruka/ 'hand', /noha/ 'leg', /liga/ 'league', /mouxa/ 'fly' ~ /ruce, noze, lize, mouše/; /cesta/ 'road', /rana/ 'wound', /vīra/ 'faith ~ /cest'e, rańe, vīře/; (b) in the loc. sing. of second-declension substantives the T ~ Ḱ alternations are productive, except *r* ~ *ř*, which like the K ~ C alternations are optional or unproductive; e.g. /plot/ 'fence', /hrat/ 'castle', /oceān/ 'ocean', /lēto/ 'summer', /stādo/ 'herd', /kino/ 'cinema' ~ /plot'e, hrad'e, oceāńe, lēt'e, stad'e, kińe/; on the other hand, /jazik/ 'tongue', /rok/ 'year', /kožīx/ 'fur coat', /vojsko/ 'army', /tābor/ 'camp', /jaro/ 'spring' ~ /jazice (jaziku), roce (po roku), kožīše (kožīxu), vojšt'e (vojsku), tāboře, jaře/; (c) in the nom. plur. of animate masculine substantives, e.g.: /ptak/ 'bird', /buldok/ 'bulldog', /pstrux/ 'trout', /hox/ 'boy', /inženīr/ 'engineer', /ad'utant/ 'adjutant', /bastart/ 'bastard', /doktor/ 'doctor' ~ /ptaci, buldoci, pstruzi, hoši, inženīři, ad'utant'i, bastard'i, doktoři/; and in the nom. plur. of animate masculine adjectives (with limitations in derivative stems), e.g. /visokī/ 'tall', /ubohī/ 'poor', /heskī/ 'pretty', /hluxī/ 'deaf', /českī/ 'Czech', /xudī/ 'poor', /dobrī/ 'good', /mlad'oučkī/ 'very young' ~ /visocī, ubozī, hescī, hlušī, čest'ī, xud'ī, dobřī, mlad'oučcī/; (d) the K ~ C alternation is productive in the loc. plur. of masculine substantives and of a few neutral substantives comprising simple stems; in derivative forms it is optional or excluded; e.g. /bok/ 'side', /klobouk/ 'hat', /monolok/ 'monologue', /hox/ 'boy', /rouxo/ 'robe', /oblāček/ 'little cloud' ~ /v bocīx, v kloboucīx, v monolozīx, v hošīx, v roušīx, v oblāčcīx/ (or /oblāčkāx/); (e) the T ~ Ḱ alternation is productive in some third-declension substantives (especially in those with the suffix *-ost'*).

In contrast with the literary language, the colloquial language has various restrictions on the occurrence of consonantal alternations. The alternations under (d) and (e) are completely or partially eliminated. Of greatest significance, however, is the elimination of consonantal alternations in the nom. plur. of personal adjectives. For while the literary language has in the plural of adjectives three genders which are expressed through different suffixes (neuter vs. non-neuter gender) and through consonantal alternations (animate vs. non-animate gender), the colloquial

language admits the opposition animate/inanimate only in the nom. vs. acc. plur. of substantives; e.g. /dobrī černox/ 'good Negro', /heskī ptak/ 'pretty bird', /českī profesor/ 'Czech professor' ~ nom. plur. /dobrī černoši, heskī ptaci, českī profesoři/ ~ acc. plur. /dobrī (or /dobrē/) černoxi, heskī (or /heskē/) ptaki, českī (or /českē/) profesori.[31] The colloquial language admits, on the other hand, alternations in the nom. plur. of substantives which do not show alternations in the literary language; e.g. /pohańi/ 'pagans', /žid'i/ 'Jews', /soused'i/ 'neighbors' (vs. literary Czech /pohanē, židē, sousedē/. In various Czech dialects, alternations occur both in the nom.-acc. plur. of substantives and of adjectives. In these dialects the alternations render the distinctions personal vs. impersonal or animate vs. inanimate gender. There are also, some dialects where the consonantal alternations redundantly differentiate the singular/plural opposition (e.g. sing. *vojāk, vojāka, vojākovi* vs. plur. *vojāci, vojācū, vojācūm*, etc.).[32] A similar differentiation of the singular/plural has been noted also in some Polish dialects.[33]

The consonantal alternations in the standard Slovak declension resemble the morphophonemic pattern of colloquial Czech, except that Slovak has the category of personal/impersonal gender, which is expressed through consonantal alternations in the substantives. The K ~ C alternation is, however, eliminated in the dat.-loc. of feminine substantives and in the loc. of masculine and neuter substantives; e.g. /ruka/ 'hand', /noha/ 'leg', /strexa/ 'thatch'; /vlak/ 'train', /prax/ 'dust'; /kongo/ 'Congo', /mlieko/ 'milk' ~ /na ruke, na nohe, na strexe, na vlaku, v praxu; v kongu, v mlieku/. The T ~ Ḱ alternation in the singular is productive. Productive also are the K ~ C and T ~ Ḱ alternations in the nom. plur. of personal substantives; e.g. /slovāk/ 'Slovak', /vlk/ 'wolf', /logik/ 'logician', /študent/ 'student', /husita/ 'Hussite', /mahomedān/ 'Moslem' ~ /slovāci, vlci, logici, študent'ḱi, husit'ḱi, mahomedāńi/. The suffix *-ia* (without stem alternation) is in many of them in free variation with the suffix *-i* (which implies consonant alternation). Adjectives do not admit

[31] The animate/inanimate opposition is optional in the colloquial language, for the nom. plural and acc. plural of substantives may overlap; cf. M. Vey, *Morphologie du Tchèque parlé*, 1946, pp. 11, 29, 31.

[32] The state of flux in the Standard Czech alternations is described statistically by L. Klimeš, "Lokal singularu a pluralu vzoru 'hrad' a 'mesto'," *Naše řeč* XXXVI (1953), 212ff; cf. also H. Rubenstein (op. cit.); for the history and dialects, cf. K. Rocher, *Gramatický rod a vývoj českých deklinací jmenných* (1934); T. Lehr-Spławinski, Z. Stieber, *Gramatyka historyczna języka czeskiego*, I, (1957), pp. 122ff.

[33] L. A. Bulaxovskij, "Voprosy indukcii grammatičeskix čisel v slavjanskoj morfologii," *Slavjanskaja Filologija, IV Meždunarodnyj S"ezd Slavistov*, 1, (1958), pp. 111, 119.

consonantal alternations, but unlike Czech, they have distinctive suffixes for the two plural genders.[34]

The youngest of the Slavic literary languages, Macedonian, has no accentual alternations; but neither is its pattern of consonantal alternations very complex. It is, nevertheless, somewhat more developed than that of neighboring Bulgarian, with which it shares the lack of a case system. Besides the K ~ C alternation, which is found in loanwords (and which is a mark of its productivity), some substantives have also the alternation of final v with s in the plural; e.g. /volk/ 'wolf', /vnuk/ 'grandchild', /belek/ 'mark', /mitink/ 'rally', /uspex/ 'success', /podatok/ 'tax', /siromaf/ 'pauper' ~ /volci, vnuci, belezi, mitinzi, uspesi, podatoci, siromasi/. And while the plural of adjectives admits no alternations in the literary language, they are somewhat more alive in the colloquial language which uses such forms as /blazi/ 'good', /druzi/ 'other'. More adjectives with alternations are found in the dialects.[35]

10. The above survey of the consonantal alternations allows us to draw the following conclusions:

Morphophonemic alternations are an integral part of the morphological structure of a language, reflecting the diversity and hierarchy of its subsystems. The verbal systems in all Slavic languages are morphophonemically more complex than the nominal systems. Within the nominal systems the declension of adjectives admits less morphophonemic alternation than the declension of substantives, and within the latter the plural admits less than the singular. The differences in morphological stratification are reflected by accentual as well as consonantal alternations. The two types of alternation have in the nominal systems an inverse and historically compensatory relationship: the productivity of one type is incompatible with the productivity of the other. Where accentual alternations are excluded, the consonantal alternations involve a large number of phonemes with a wide range of distribution, and carry distinctive functions; whereas in systems which utilize accentual alternations, all of these characteristics are rare.

It could, perhaps, be argued that the role of consonantal alternations may be as limited in languages without accentual alternations as in those

[34] Cf. E. Pauliny, J. Štolc, J. Ružička, *Slovenská Grammatika* (1955), pp. 134ff. The history of the declension (with scattered references to the consonantal alternations) is treated in J. Stanislav's *Dejiny slovenského jazyka* (1958), 2; cf. esp. pp. 58–68, 121–126, 145–150.

[35] Cf. B. Koneski, *Gramatika na makedonskiot jezik* (1952), 2, p. 73, and A. Vaillant (op. cit.), 2, p. 519

with them. The Macedonian and Bulgarian nominal systems are a case in point. However, several factors must be considered. First, the accentual alternations in Bulgarian play a far less significant role than they do in other languages which have them (e.g. Russian, Slovenian, Serbo-Croatian). Second, geographic and historical closeness of languages entail phenomena of diffusion and of influence on the part of older literary languages; in the case of Macedonian, on the part of Bulgarian and of Serbo-Croatian. And third, the consonantal alternations could have been reduced in Macedonian prior to the loss of accentual alternations; stress is still phonemic in some Macedonian dialects. With the loss of phonemic stress, both types of alternation were affected. But the most important point is this: in a typological classification, as in any other, it is necessary to establish the polar terms which are clearly opposed to each other, and then proceed to the transitional types. And it is obvious that none of the languages with accentual alternations have such a rich system of consonantal alternations as Polish, Lusatian or Czech, whereas none of the languages without accentual alternations lack altogether consonantal alternations, as do Russian or Slovenian.

Department of Slavic Languages
Indiana University
Bloomington, Ind.

WILLIAM R. SCHMALSTIEG

A Note on Slavic Verbs of the Type *zějǫ* : *zĭjati*

In Slavic there is a class of verbs which is characterized by an *-e-* grade root vocalism in the present stem as opposed to a zero grade (traditionally called reduced grade) in the aorist stem.[1] For example, we find the following pairs:

Present (1st sg.)	Infinitive (aorist stem)[2]
berǫ	bĭrati 'to take'
derǫ	dĭrati 'to tear'
ženǫ	gŭnati 'to chase'
zovǫ (< *zev-ǫ)	zŭvati 'to call'

There are certain verbs in this category with a stem final *-j-*.[3] These differ from the others in that they have Slavic *-ě-* (< IE*-ē-) in the present stem. Thus we find:

Present (1st sg.)	Infinitive (aorist stem)
zějǫ	zĭjati 'to yawn'
směję (sę)	smĭjati (sę) 'to laugh'
lějǫ	lĭjati 'to pour'

[1] Chr. S. Stang, *Das slavische und baltische Verbum* (Oslo, 1942), p. 90.

[2] In general there is a problem in Slavic as to why we find the vocalic reflex of the sonants, i.e. *-ĭr-*, *-ĭl-*, etc. (< IE*-*ṛ*-, *-*ḷ*-, etc.) in pre-vocalic position at morphological boundaries, e.g. *bĭr-a-ti* instead of **br-a-ti*. In *Language* XXXV (1959), 16–17 I suggested that this might be the result of generalization of the vocalic plus consonantal form of the sonant which is to be expected in initial position according to Edgerton's formulation of Sievers' law. Whatever the origin may be, there is certainly a morphological parallelism between *-ĭr-*, *-ĭl-*, *-ĭm-*, *-ĭn-* (*-ŭn-* in the case of *gŭnati*) on the one hand and *-ĭj-* and *-ŭv-* on the other. I shall term all of the above reflexes of the sonants the *vocalic zero grade*, including even *-ĭj-* and *-ŭv-* although strictly speaking here we have vocalic plus consonantal reflexes.

[3] Stang, p. 92 and Nicolaus van Wijk, *Istorija staroslavjanskogo jazyka* (Moscow, 1957), p. 341 = Russian translation of his *Geschichte der altkirchenslavischen Sprache*.

Stang says that he can give no satisfactory explanation for the existence of the lengthened grade *-ě- in the present stem.[4]

For the verb *zějǫ* we can refer to Vasmer's statement that this form can only reflect an IE**ĝhēi̯ō*.[5] If, however, we accept Diver's theory that there existed in IE a phoneme **-Hᵘ-*, a laryngeal with concomitant palatal articulation,[6] then we can change Vasmer's reconstruction to **ĝheHᵘ-*, thereby explaining both the etymological *-ě- and the fact that *-j-* belongs to the root and not the suffix.

Since in the zero grade both the phonemes IE**-j-* and **-Hᵘ-* are represented by the same phoneme (cf. Skt. *sy-a-ti* < **sHᵘ-e-ti*),[7] it seems reasonable to suppose that the phonemic complex *-ĭj-* could be a transformation of either IE**-j-* (as is generally supposed) or of **-Hᵘ-*. Thus the cognates Lat. *hiāre* and O.C.S. *zĭjati* can be traced back to a form of **ĝ̑hHᵘ-* with vocalic zero grade (perhaps the generalization of a Sievers'-law variant with vocalic plus consonantal function, following a heavy syllable or in initial position). According to traditional explanations the stem for 'to yawn' is a heavy base and therefore the Lith. *žioti* /ž̑oti̯/ seems to pose a problem. We suggest that the form *žioti* be traced to **ĝhHᵘ-* also, but that in this case we have a generalization of the consonantal reflex of **-Hᵘ-* (i.e. only the palatal articulation) rather than vocalic plus consonantal as with O.C.S. *zĭjati*, etc. The Skt. cognate *vi-hāyas* 'atmosphere' reflects the full grade of the root **ĝheHᵘ-* as does the Slavic present stem *zějǫ*.

On the other hand the verb *směję* (*sę*) 'I laugh' has cognates which do not presuppose a laryngeal, cf. Gk. *meidáō* 'I smile', Tokharian A *smimāṃ* 'smiling', Lat. *mīrus* 'marvelous', Skt. *smayate* 'he smiles', but zero grade *smitas*.

The form *směję* (*sę*) is to be explained by analogy with *zějǫ*. An original **smejǫ* with a short *-e-* would have given Prim. Slavic **smĭjǫ*, since IE**-ej-* becomes *-ĭj-* in pre-vocalic position.[8] In other words the **-e-* grade would have been homonymous with the vocalic zero grade. In other verbs of this category a vocalic zero grade contrasts with the *-e-* grade, cf. *bĭrati*: *berǫ*, etc., but with a stem final *-j-* no such contrast is possible

[4] Op. cit., p. 92. An explanation sometimes advanced is that given by André Vaillant (*Grammaire comparée des langues slaves*, Paris, 1950, Vol. I, p. 99) that the group **-ejj-* became *-ěj-* in Slavic. This does not seem likely to us because the IE diphthong **-ej-* passes to Slavic *-i-* before every other consonant, e.g. Lith. *véidas* 'face', Slavic *vidŭ* 'aspect'.

[5] Max Vasmer, *Russisches etymologisches Wörterbuch* (Heidelberg, 1953), p. 457.

[6] William Diver, "Palatal Quality and Vocalic Length in Indo-European," *Word* XV (1959), 110–122.

[7] Diver, *op. cit.*, p. 119.

[8] Antoine Meillet, *Le slave commun* (Paris, ₂1934), p. 48.

because of the phonemic structure, i.e. **-ej-* did not exist in Prim. Slavic. As a result of this we suppose that the contrast *-ěj-*: *-ĭj-* (as in *zějǫ*: *zĭjati*) became the morphological replacement of **-ej-*: *-ĭj-*.

Slavic *lějǫ*: *lĭjati* can be explained either as a root with **-Hᵛ-* or as analogical to *zějo*: *zĭjati*. The acute pitch stress of the cognate infinitives Lith. *lýti* 'to rain' and S.Cr. *lîti* 'to pour' speaks in favor of an original laryngeal. Likewise the O. Lith. pres. conj. *lėju* which reflects IE**lēi̯ō*, according to Pokorny,[9] suggests a laryngeal. The Prim. Baltic *lej-tei*, reflected by Lith. *líeti* (with East Baltic **-ei-* > *-ie-*) is from substitution of the prevocalic form **lēj-* in position before a consonant with the Common Baltic shortening of long dipthongs in tautosyllabic position. The modern present conjugation *líeju*, etc., was then reformed on the basis of the infinitive.

Church Slavic *prějǫ*: *prĭjati* 'to have care for' has the cognates Goth. *frijōn* 'to love', Skt. *príyate* 'he is satisfied' and Pokorny cites as basic forms *prāi*, *prəi*, so we may guess that this originally had a laryngeal root.[10]

Department of Languages
Lafayette College
Easton, Penna.

[9] Julius Pokorny, *Indogermanisches etymologisches Wörterbuch* (Bern, 1949ff.), p. 664.
[10] *Op. cit.*, p. 844.

MICHAEL RIFFATERRE

Stylistic Context

1.1. The definition of literary style as a departure from the linguistic norm raises difficulties of application in style analysis. In a previous paper[1] I therefore proposed to replace the notion of overall norm with that of stylistic context, and to study stylistic devices (hereinafter *SD*) in relation to this context. The context, by definition inseparable from the SD, (1) is automatically relevant (which is not necessarily true of the norm); (2) is immediately accessible because it is encoded, so that we need not rely on an elusive and subjective *Sprachgefühl*; (3) is variable and constitutes a series of contrasts to the successive SDs. Only this variability can explain why a linguistic unit acquires, changes, or loses it stylistic effect according to position, why every departure from the norm is not necessarily a fact of style, and why style effects occur without abnormality.[2]

1.2. The meaning of a message can be received with minimal decoding. More is required if the writer wants to force upon his reader's attention certain formal features to which he attaches special importance (e.g. esthetic intent). But what permits minimal decoding is that it is possible with variable accuracy to predict, from part of a sequence, the features which follow. Control of the decoding, then, is achieved through low predictability.[3] We can therefore define the stylistic context as a pattern broken by an unpredictable element (this contrasting factor being the SD).

Style is not a string of SDs, but of binary oppositions whose poles (context/SD) cannot be separated. Far from concentrating on the SDs just because they are the salient, easily classified features, style analysis (and later, esthetic metanalysis) must devote as much attention to the unmarked

[1] In my "Criteria for Style Analysis" (hereinafter *Criteria*), *Word* XV (1959), 154–174, esp. §§3.0–3.2.4; see also my review of S. Ullmann's *Style in the French Novel*, in *Word* XV (1959), 404–413, esp. pp. 407–409.

[2] On this variability of effects, or *polyvalence*, see my review cited footnote 1, page 409. Similarly, L. Spitzer, after comparing—with a very different approach—more styles than any stylistician ever had, concluded that the stylistic sign is "empty" (see, for example, *Modern Language Quarterly*, XIX [1958], 235, footnote 8).

[3] See *Criteria*, §§1.1–1.2.

elements of the oppositions. This paper, mostly orientational, seeks to give as simple and complete an account as possible of the various aspects of the context.

1.3. Two categories of facts require an expansion of the above definition: *A*. Certain literary pieces are so short that they seem not to have "time" to form the contextual pattern which permits contrasts: e.g. *haiku*, monostich, prose aphorism, motto, epigraph, etc.;[4] other similar cases: the beginning of literary utterances and the clusters of SDs, or *convergences*,[5] where the accumulation of SDs does not seem to allow enough context between them to prevent saturation (and yet saturation does not occur). *B*. The context could be represented by a linear segment, following the decoding process, therefore oriented in the same direction as the sentences. The existence of this orientation is implied by the process of reading, by word order devices, and by rhythms. However, in short texts, like those mentioned above, the SDs can be recorded at a glance, or the end of the passage can be perceived, in advance of actual reading, by peripheral vision.[6] Furthermore, even in texts of ordinary length, the reader can always reread what stopped him in his first casual decoding.[7] In all these cases, we may wonder whether the effect of surprise does not dwindle and whether orientation is not intermittent. These limitations to the scope of the definition can be removed if we introduce a distinction between context inside and outside of SD, that is, between the context which creates

[4] The size limits of this type are very narrow: a poem hardly longer than a monostich, such as the Latin elegiac distich, would not fall in the same category, since the complexity of the metric pattern makes up for its brevity. Quotations from other authors must also be excluded, because they are provided with a new context. Dictionaries of quotations, on the contrary, should be included: they are meta-texts only for the style analysis of a literary work as a whole, but isolated, without consideration of the compiler's reasons for excerpting them, they may be studied to determine to what extent a context can be cut off without destruction of the stylistic effect.

[5] On this phenomenon and its name, see *Criteria* §3.2.3. It is encoded in the written text and must not be confused with the concurrence of several sets of signals, such as can be observed in the *performance* of certain oral literary forms (see T. A. Sebeok, "Folksong Viewed as Code and Message," *Anthropos* LIV [1959], 141–153, esp. 141–42).

[6] On perception in reading (word recognition, peripheral vision, etc.), see the handy summary of M. A. Tinker "Visual Apprehension and Perception in Reading," *Psychological Bulletin* XXVI (1929), 223–240; the yearly summaries of "Reading Investigation" by Wm. S. Gray in the *Journal of Educational Research*.

[7] It may be an almost unconscious re-reading by peripheral vision; or a conscious search for the solution of some decoding problem: this search entails several simultaneous Markov processes working forward and backward (see a good example in A. A. Hill, "An Analysis of *The Windhover*: An Experiment in Structural Method," *Publications of the Modern Language Association* LXX (1955), 968–978, esp. 975–976).

the opposition constituting the SD, and the context which modifies this opposition by reinforcing or weakening it. This solution makes it unnecessary to assume the existence of two styles of different natures, whose "cohabitation" within the same text would be hard to explain, and enables us to outline a hierarchy of structures without gaps or incompatibilities, wherein higher levels of structure can be stated in terms of the units of the preceding level. I shall call the first type *microcontext* (rather than segmental) and the second *macrocontext*.

2.1. Definition of the microcontext. (The term *context* here is not taken in its common meaning; it is considered independently from any influence of the macrocontext, although such an influence always exists, except under the conditions described in 1.3.*A*; this special use of *context* may be awkward, but does not seem much more objectionable than, say, the coexistence of a linguistic and a common sense *redundancy*). Let there be in a (literary) sequence a group of features bound together on one or more levels of the linguistic system by a structural and semantic relationship. If the group has a stylistic effect, its stimulus consists of less predictable elements encoded in one or more constituents. The microcontext consists of the other constituents which remain unmarked; contrast is created in opposition to these constituents (the reader perceives the degree of unpredictability in relation to them).[8] The group as a whole (context + contrast) forms the SD. The essential characteristics of the microcontext are: (1) it has a structural function as a pole of a binary opposition, and consequently: (2) it has no effect without the other pole; (3) it is spatially limited by its relationship to that pole (in other words, it does not encompass elements irrelevant to the opposition and may be limited to one linguistic unit). Its constituents can be multiple, discontinuous (e.g., the disjunct group in a disjunction), or simultaneous (e.g., the unchanged part of a renewed cliché, the blended words in a portmanteau word).

It should be possible to build a (stylistic) grammar of the conditions under which contrasts occur; some of their aspects are obvious enough to permit the establishment of relative measurements; for example, the effectiveness of the contrast is in direct proportion to its degree of unpredictability, that is, to the degree of predictability allowed by the internal context.[9]

[8] It is obvious that I do not start my analysis from this unmarked element (at the present stage, this would amount to falling back upon the concept of norm), but from the contrast immediately perceived by the reader; for a discussion of this approach, see *Criteria*, Chap. 2.

[9] Like this paper, the examples given are orientational: they are introduced as further clarification; but no measurement will be attempted in the present study. Once again, I

Ex. (1*a*) Corneille, *Cid*, I.1273: *Cette obscure clarté qui tombe des étoiles* (semantic contrast; both *obscure* and *clarté*, as separate units, have a high determinacy, since the topic is the military use of darkness; but as constituents of a single structure, they form the least probable unit; *clarté* stands in contrast with the context *obscure*; *obscure clarté* is the SD apprehensible as a unit, which is demonstrated by the frequent quotation since then of *obscure clarté* as an isolated phrase).[10]

(1*b*) Use of certain prepositions for others, different in function, such as French *à* for *en*, *sur*, *sous*; or of conjunctions stripped of their function, such as *car* used where no causative construction is in order (both devices were characteristic of the French Symbolist school);[11] such SDs are strong because they affect the very structure of the grammatical system.

(1*c*) Pope, *Rape of the Lock*, II, 105–109: *Whether the Nymph shall* (. . .) *stain her Honour, or her new Brocade,* (. . .) *Or lose her Heart, or Necklace, at a Ball.* (syllepsis:[12] the metaphorical meaning of the verb in the two contexts—*stain her honour, lose her heart*—makes the shift to its ordinary meaning unpredictable; this and the resulting forced parallels *honour/brocade*, *heart/necklace* impose maximal decoding).

In each case the context compensates for its brevity by narrowly limiting the reader's choice of predictions.

2.2. Microcontexts in short sequences. Thus it is clear that microcontexts make style possible even in a short sequence. It is enough that the sequence contain one or more SDs or a convergence of SDs; these exclude casual decoding. Ex. (2) Emily Dickinson, *Bolts of Melody*, Fragment 635: *Or fame erect her siteless citadel* (a case of self-sufficient style, since this isolated line was found noted on the flap of an envelope, indicating the writer's desire to preserve a form rather than a content. Stylistically, *fame erect* and *siteless citadel* contain each a contrast which in both cases impose metaphorical interpretations; *citadel* is more predictable in the frame *erect X*, but the contrast is immediately reestablished by *siteless*; furthermore *siteless* is expressive because it is a nonce-word in opposition to

am trying only to provide a coherent organization of materials, using nothing but relevant features.

[10] Note that *clair-obscur* does not appear in French until 40 years later, and then only in its Italian form and in technical parlance: there was no association susceptible of weakening the SD.

[11] See L. Spitzer, *Aufsätze zur romanischen Syntax und Stilistik* (1918), pp. 289ff.

[12] This example is borrowed from Justin O'Brien, "Proust's Use of Syllepsis," *PMLA* LXIX (1954), 741–752; this article can be used as a monograph on syllepsis in general. On zeugma, closely related to syllepsis, see G. O. Rees, *Français Moderne*, XXII (1954), 287–295; R. Le Bidois, *ibid.*, XXIV (1956), 81–89, 259–270.

words of the common code and because it forms a contrast—non-existence versus existence—to both *erect* and *citadel*).

In short sequences, there is no macrocontext to heighten the effect of a SD (§3.2.); on the other hand, none diminishes it either (§3.3.).

2.3. Maximal decoding. There remains the question of the surprise effect, which must be distinguished from the unpredictability which causes it. The latter defines statistically a certain mode of occurrence. The SD presenting this occurrence may well be perceived at a glance or anticipated by peripheral reading: this will undoubtedly lessen the surprise effect. But it cannot diminish the sense of abnormality (abnormality *hic et nunc*, not in relation to the awareness of the norm) which results in increased attention, i.e. maximal decoding.

Rereading, for example, may resolve a semantic ambiguity, but it does not modify the order of the contextual elements; we do not read backwards, we jump back and then move again in the usual direction. Our memory, now working forward, informs us of what is to follow, and in fact destroys surprise; but, by the same token, we again pass through the forms whose unpredictable arrangement had necessitated rereading, and this process imprints them more deeply than did the initial surprise.[13]

On the other hand, predictability in itself is not an absolute, except within stereotypes. This permanent uncertainty, and the margin which must exist between statistical and perceived unpredictabilities would reduce surprise (although it remains basic in devices such as: syllepsis; anacoluthon—breaking off of the sentence when it has reached the point where readers can guess the end—; adve-b where an adjective has a higher probability). Perhaps this is the reason why surprise is frequently supplemented by the effect of the macrocontext (see §3.2.). Within the microcontext, however, the fact that the reader is forced into a more thorough decoding is enough to account for the functioning of the SD, even if surprise is entirely absent. Such maximal decoding is dictated, for instance, by disjunctive word-order: the interruption of the sentence compels the reader to make an effort of retention and attention until the sequence is resumed. Ex. (3) E. M. Forster, *Collected Tales* ("Other Kingdom"), p. 82: *My name is Harcourt Worters—not a well-known name if you go outside the*

[13] Cf. the related question of the duration of stylistic effects through successive "performances" of the text (see *Criteria* 1.3.0–1.3.1; Wellek and Warren, *Theory of Literature*, 1956, Chap. XII). The conditions of unpredictability are maintained in the "frozen" structure of the message; thus successive performers will be exposed to the same phenomenon. For a reader "performing" repeatedly, memorization will cancel unpredictability, but will preserve the formal characteristics of the text. Whether

City and my own country, but a name which|, where it is known|, carries|, I flatter myself|, some weight. (The disjunctions would not suffice to nullify the predictability of the word sequence, especially after *carries* which introduces a stereotype. What brings out the speaker's pompous, self-righteous vanity, what produces the contrast is the suspension of the sentence after *which*.)

Effects of surprise can be combined with those of maximal decoding (e.g. when, after the disjunction, the sentence is not resumed in the expected or most probable fashion, but with a different construction or vocabulary).

3.0. Definition of the macrocontext. The macrocontext is that part of the literary message which precedes the SD and which is exterior to it. (It is this type of context which is closest to *context* in its everyday meaning.) Since its decoding supposes a spatial orientation, we see that the phrase *oriented linear segment* in our preliminary definition applies to this context: it could not have applied to the microcontext. This characteristic permits the storing up of information which will ultimately modify the effect of the compound (microcontext/contrast) SD.

The terminus of the context is known by definition; its starting point is not. Insofar as its effect on the SD is known through the reader's reactions, the point of the earliest perception of context might be said to vary in function of the reader's attention and memory, that is, his ability to recognize formal similarities and dissimilarities. On the other hand, if the SD is a rupture effect, it is tempting to conclude that the context reaches back to the last occurrence of an identical or similar SD. Such a solution, however, would not account for cases where the context precedes a hapax phenomenon, and cases where the SD is a *tic* of the author's speech (in this instance, a given form is surprising and easily noticeable because it is characterized by recurrences in semantic or situational contexts, where it has no conditioning other than psychological; a *tic* then is recalled over lengths of text whose extent and variety far exceed what could be reasonably attributed to the context as a functional unit). The starting point cannot be seen in *any* SD preceding the one under study, since, as we shall see (§3.1), macrocontexts may contain SDs. It cannot be said that any linguistic boundary, such as the limit of a sentence, marks *de facto* the starting point of a context, since stylistic effects may overlap linguistic

memorized or unexpected, style remains an active factor. As for literary imitation, it renews the SD by transferring it to a new context (e.g. Vergil using Homeric devices): this last process is no different from the use of a cliché in a new environment.

units (although, in fact, the context often starts with a paragraph or other punctuation).[14]

Probably we should here admit a certain area of indeterminacy, perhaps a permanent one if we consider that any notion of limit in style analysis is only approximate because of the reader's variable threshold of perception. The starting point can then be identified more closely if we use the very definition of the context, that is, the notion of pattern. We should not posit a pattern homologous to the SD: the homology may indeed exist (e.g. in French non-interrogative sentences, a pattern of clauses presenting verb-subject order may prepare a contrast with subject-verb order, which elsewhere would be "normal"); it is not of course a necessary condition, since a pattern can be disrupted by elements foreign to it (e.g. foreign language quotations or nonsense utterance; sometimes zero elements, such as suspension points). But it is safe to substitute continuity for homology and to say that the context begins at the point where the reader perceives the existence of any continuous pattern.[15]

3.1. Types of macrocontexts. Two types seem to me to cover all possibilities: (*A*) *Context → SD → Context*. This type is characterized by the resumption after the SD of the contextual pattern which prepared it. The most frequent example of this type is the injection into the context of a word foreign to the code used (borrowing, archaism, neologism).[16] Ex. (4) G. B. Shaw, *Getting Married*, Preface (*Selected Plays*, IV, 329):

[14] *A.* Note that in this respect microcontext can be relatively long: even a repetition, for instance, can be the immediate constituent of a SD; the longer the repetitive pattern lasts before being broken, the stronger the rupture effect—*B.* Of course we must distinguish: (1) the repetition as SD, that is, a contrast-creating structure which consists of the rhythmic or arhythmic recurrence of an element (e.g. a verbal ending); (2) the repetition as microcontext; (3) the repetition of SDs, that is, the recurrence of the same stylistic structures with various contents, this being a case of loose terminology, the normative metalanguage of traditional literary criticism—*C.* I use here (and elsewhere) the verb *prepare*. I mean it to describe a context building toward a SD (e.g. in a period, free rhythms preparing the contrasting clausula, or a long protasis preparing the antithetic climax of a short apodosis); I never mean it as a statement of the writer's intentions.

[15] *Continuity* should not be taken as an assumption about the author's psychological processes. Without doubt the pattern corresponds to a certain continuity in the writer's mind (true even of *écriture automatique* in French Surrealistic writings: this device was in deliberate opposition to the continuity produced by reflection, but it aimed nevertheless at recording the continuity of subconscious mental activity), but it does not *depict* it in the form of a sequence of words; to believe so would be a metalinguistic confusion (cf. the "chronological" interpretation of Latin word-order in sentences like *urbem captam hostes diripuerunt*).

[16] See Joshua Whatmough, *Poetic, Scientific and other Forms of Discourse* (1956), pp. 105–107.

Poor Mr. Pecksniff (. . .) *is represented as a criminal instead of as a*/*very typical English*/*paterfamilias*/*keeping a roof over the head of himself and his daughters*. (The borrowing *paterfamilias* is unpredictable and herein lies the effect of the device, whose immediate constituents are *very typical English* and *paterfamilias*. This effect is oriented by contrast with the context, which emphasizes ordinariness and its ridiculous side [proper name, cliché *keeping a roof over*]; the orientation could be different if Shaw, instead of a borrowing already in use and stylistically marked, had used his own nonce-word; the mechanism of the device would still have been based on surprise. Following this device, the sentence continues with its ordinary vocabulary and the resumption of the interrupted line helps make the SD appear *relatively* abnormal.)

(*B*) *Context* → *SD starting new context* → *SD*. The SD generates a series of SDs of the same type (e.g. after an SD by archaism, proliferation of archaisms); the resulting saturation causes these SDs to lose their contrast value, destroys their ability to stress a particular point of the utterance, and reduces them to components of a new context; this context in turn will permit new contrasts. Ex. (5) Tacitus, *Vita Agricolae*, XXXVII, 1–6; too long to be quoted here, this narrative of a Roman victory over the Britons contains the following sequence of tenses: (a) a series in the pluperfect, imperfect and perfect/ (b) a series in the infinitive of narration/ (c) one historical present/ (d) a series similar to (a). This sequence corresponds to dramatic shifts in the action: (a) the confusion of battle/ (b) the sudden victory and general slaughter/ (c) the last surge of resistance/ (d) the mopping-up operation. The sequence, translated in terms of stylistic opposition is: (a) *context*/ (b) *SD composed of a group of three infinitives* (supported and unified by a convergence of expressive meanings and asyndeton: *sequi, vulnerare, capere*) *which set off a context made of infinitives* (this time without convergence and spaced by long phrases)/ (c) *SD*/ (d) *resumption of the initial context*.[17]

There might be an extreme case in which the SD, instead of creating a new context, would function as a context for another device following immediately. Instances of this type are by no means clearly defined, because saturation tends to cancel the contrast.

3.2. The emphatic function of macrocontexts. The context reinforces the stylistic effect of the device. It amplifies the contrast established by the

[17] We find many similar instances in the interplay of *narrative*/*direct discourse*/ *indirect discourse*/ *free indirect discourse*; see S. Ullmann, *Style in the French Novel*, pp. 94–120; bibliography of this device, in A. G. Landry, *Represented Discourse in . . . Mauriac* (1953), pp. 76–79.

microcontext. This is probably the most frequent effect in style, and it is obtained in both types *A* and *B*; indeed it is the only effect possible with type *A*. Without attempting to cover all modes of emphasis, I shall indicate two opposite aspects of its function which may serve to point to fruitful analytic procedures.

The contrast of the SD may be prepared by a stylistically unmarked context: e.g. prevalent use of the favorite sentence-type will constitute a stylistic zero-degree by opposition to which SDs will stand out; "casual" speech may have the same effect. However "casualness" has yet to be defined; linguistic analysis permits us to approach a definition: but a relevant definition should cover only the forms "casualness" assumes in *written* speech; moreover we must still determine to what extent the reader is conscious of or reacts to what linguistic analysis describes (in other words, *stylistic* analysis should be limited to what is accessible to the reader).[18] In any case, the existence of such "blank" contexts can be experimentally confirmed by examining writers' successive drafts: these often show deletions intended to create a vacuum in the midst of which the SD is conspicuous.[19] Such deletions demonstrate that context is as important a part of style as the SD. Another aspect of this fact is the role of context as a mechanism to space out the SDs; aligned in too close succession, they may cancel one another (see 3.3); their effect may vary in proportion to their proximity (for example, a common artefact of dramatic dialogue, the repeated phrase peculiar to a given character, may be a tool of psychological realism; but if repeated more often, that is if the contextual spacing between occurrences is shortened, it becomes a comic factor).

A more frequent form of the emphatic function, however, is anticipatory: the context prefigures one of the immediate constituents of the SD. Chiasmus is a case in point. Ex. (6): given the SD (*microcontext*) *sleep eternal*/ (*contrast*) *in an eternal night*/, if the macrocontext contains a

[18] It could be said that this type of context is close to the linguistic norm. The departure, however, is not from the "consciousness of the norm," but from a written background.

[19] Editors classify these deletions indiscriminately under the head of *concision*; but this involves the introduction of the idea of simplicity, which is an esthetic criterion hard to define and irrelevant in an inquiry as to the *existence* of style; the deletion amounts to an increase, objectively measurable, in the distance between SDs. E.g. Balzac, who is far from aiming at simplicity, deletes many emphatic devices (*tout* modifying adjectives, *c'est . . . qui*/ *que . . .*, etc.). These features, frequent in speech, are dropped significantly from written conversations: without intonation, they would only be padding, and their omission permits a rapid succession of sudden contrasts more suggestive of the vividness of actual speech than a total recording would be. See Mario Roques, "Manuscrit et éditions du *Père Goriot*," *Revue universitaire* XIV (1905), 34–42, 71–76, 178–183.

noun-adjective pattern (a pattern conventionally characteristic of poetic utterances), the parallel structure links macrocontexts and microcontexts; e.g. Swinburne, *The Garden of Proserpine*, 80ff.: *Then star nor sun shall waken/ Nor any change of light;/ Nor sound of waters shaken,/ Nor any sound or sight;/ Nor wintry leaves nor vernal,/ Nor days nor things diurnal;/ Only the sleep eternal/ In an eternal night.* (Here there is an immediate contact between the contrasting feature and what precedes it, and the stylistic effect is further heightened by the cumulative sequence of the pattern, which reinforces its homologue within the SD.)[20]

3.3. The leveling function of macrocontexts. By this I mean the destruction of the contrasts, the result of which is that the SDs sink to the level of the context and are no longer distinguished from it (e.g. the loss of effect a neologism suffers when surrounded by too many nonce-words). This leveling function is possible because the macrocontext is a cumulative process. The function has two possible sources: it may be the consequence of shorter spaces between SDs—greater proximity prevents the formation of patterns that would prepare contrasts; or else it may result from a type B context, in which a given SD can expand into a pattern which in turn creates a bipolar opposition to another SD.

The intensity of this phenomenon is in direct proportion to the frequency of occurrence of devices belonging to the same class.[21] We can speak of the SD as a contaminating element which modifies the role of the context as the "unmarked" member of an opposition.

3.4. The general implications of the above may eventually program future style analysis: since devices can be reduced to the level of the context, any concept of an intrinsic stylistic value is void (e.g. the expres-

[20] Needless to say, this is a very partial analysis. It does not take into account rhyme and rhythm, or concurrent and overlapping SDs such as (*context*) *nor . . ./ only . . .*, which in turn contains *nor any . . ./nor . . .*.

[21] If the devices were identical in form and content, we would have *one* SD: repetition (cf. footnote 14). The reasons for these high frequencies range from unconscious or conscious self-imitation (on stylistic contagion, see L. Spitzer, *Linguistics and Literary History*, p. 191, n. 46; J. Marouzeau, *Revue des études latines* XIV (1936), 58–64; XXVI (1948), 105–108) to a reaction against normative precepts forbidding repetitions. We shall have to inquire whether there are SDs whose microcontext makes them immune to saturation (e.g. disjunctive word-order, because it suspends decoding altogether) and we shall need statistics, based exclusively on stylistically relevant data. This is in synchrony a problem similar to that of the duration of SDs in diachrony (i.e. resistance to wearing out through repeated performances and "traditionalization"; see, for example, M. Riffaterre, "La durée de la valeur stylistique du néologisme," *Romanic Review* XLIV (1953), 282–89).

sivity "inherent to" iterative verbal forms, superlatives, etc.). In its place, the leveling function is distributional in nature and therefore measurable. Consequently it will be possible to define conditions of occurrence. Unlike those already known through the direct application of statistics to the linguistic message (these sets can define the literary utterances but cannot differentiate the linguistic material and the stylistic use thereof), we seek conditions relevant only to what is *perceptible as style*.[22]

As an analytic criterion, the concept of context has varied applications. When literary language (or more specialized systems such as verse) is treated as a restricted norm, that is, with more limitations than the overall linguistic norm, there is the danger of seeing style only as a further restriction and to evaluate it by mere subtraction (e.g. by identifying "unpoetic"[23] words). But if literary language is analyzed as a sequence of contexts, then proper emphasis will fall on the enrichment furnished by new stylistic possibilities: these conventionally structured contexts make SDs out of features which would have no contrasting value in a "casual" context or which would saturate it and thereby alter its nature; a conventional frame permits stronger effects without such disfiguring modifications.[24] The leveling function presents the advantage of accounting for what I have called elsewhere *nonce-standards*, the fragmentary language systems which the author uses to suggest the speech of one of his characters or to parody a style. These can be seen as special contexts developed by saturation from SDs originally borrowed by the author from the subject of his evocation and set against the background of his own style.[25] Study of these particular

[22] The *mot juste*, for example, may be defined by its fitness, whose measure in language is statistical (J. Whatmough, *Poetic . . . Discourse*, pp. 119–121). But its stylistic relevance does not lie in its fitness, which is defined in relation to the linguistic structure; it lies in the variability of its effects in relation to different contexts (and to different poetic dogmas). The *mot juste* may be a SD, or it may create a context so "fitting" that vague, "unfitting" words will become devices by contrast.

[23] For example, in Latin, see Bertil Axelson, *Unpoetische Wörter*, Lund, 1945.

[24] It is significant that poetic language frequently uses *Umgangssprache* forms: they would have no effect in their usual context; they have an effect in the contrived setting of meter. Pointing in this direction are observations such as Meillet's that "la poésie (latine) n'a pas le purisme de la prose" (*Esquisse d'une histoire de la langue latine*, 1938, p. 220); prose, unprotected by conventional forms, was not able to use "ordinary" words without losing its literary character. (Cf. frequent remarks in poetic glossaries that certain words belong equally to the "poetic and familiar" vocabularies: clearly they are SDs in the former, contextual in the latter). See in French the use of substandard speech forms *l'*, *m'*, *j'*, *n'* (*le, me, je, ne*) in Jules Laforgue's poems. Normative conditioning of the literary message, which linguistic analysis bypasses, should become apparent (see L. Kukenheim "Réflexions non-structuralistes," *Neophilologus*, 1955, 161ff.).

[25] See *Criteria*, p. 169, n. 24 and footnote 12. (These must not be confused with

cases should give a purely linguistic basis to the description of literary realism and of ironic styles, for which external criteria (psychological, normative, etc.) are generally employed. In another direction, we could ask whether the leveling at a given point can become irreversible attrition or whether positional factors suffice to renew stylistic effects; the answers to such questions may ultimately permit the elaboration of a diachronic stylistics.

Department of Romance Philology and French
Columbia University
New York 27, New York

situational norms, predetermined by the topic.) For example: elements suggesting (not transcribing) either foreign pronunciations (e.g. speech of German characters in Balzac, of Jews on the English Victorian stage), or idiolectal features used for psychological characterization (e.g. Mr. Jingle's way of speaking in the *Pickwick Papers*), or social attitudes (e.g. parodic use of stereotypes; see M. Riffaterre "Sur un singulier d'André Gide. Contribution à l'étude des clichés," *Français Moderne* XXIII [1955], 39ff.), or other authors' styles (e.g. Proust's pastiches; imitations of the Bible; for the latter even one feature may suffice, see L. Spitzer, *Linguistics and Literary History*, p. 150; Uriel Weinreich, "On the Cultural History of Yiddish Rime," *Essays in Jewish Life and Thought*, 1959, p. 423ff., esp. 434–435; on a much more complex mimetic standard, see Y. Le Hir, *Lamennais écrivain*, 1948, pp. 241–332). A good example of SDs which create a nonce-standard is the hyphenated neologisms Carlyle employs in *Sartor resartus* to suggest "German-philosophical" style.

ALVAR ELLEGÅRD

Estimating Vocabulary Size

How many words are there in a language like English? What is the normal vocabulary of an educated adult native speaker? How many words should a student know at various stages of learning a foreign language? What do we mean by saying that a writer has a rich vocabulary? How many words does an ordinary novel contain? How often should we expect a dictionary containing 10,000 (or 50,000, or 100,000) entries to fail to give us any lead in the course of reading an ordinary book of 300 pages?

These are some of the questions concerning vocabulary size which are of interest to linguists, psychologists, teachers, lexicographers (including machine translators), and perhaps even to ordinary people. The present paper is an attempt to provide some of the basic facts and arguments which are needed to answer them.[1]

1. The first problem to be treated is this: Given an average text consisting of *W running words*, how many *different words V* will it contain, and how many times will each word in that vocabulary occur?[2]

It might seem that a proper answer to that question would involve a tremendous amount of sheer mechanical counting of a large number of text samples of different sizes. For instance, we might select, say, ten samples of 10,000 words, another ten of 20,000, and so on. On the basis of such counts, a diagram could be drawn up, from which the average number of different words to be expected in a text mass of any size could be read off.

Some data which might be included in such a diagram are already in existence: a few will be discussed below. But they are not nearly sufficient, and to obtain what is needed would be a formidable and costly undertaking, even with the aid of electronic data processing machines.[3] It is

[1] Some of these problems have been treated in G. U. Yule's classical *The Statistical Study of Literary Vocabulary*, Cambridge, 1944: and more recently in Pierre Guiraud, *Les Caractères Statistiques du Vocabulaire*, Paris, 1954.

[2] The symbols used are Yule's. *W* stands for "words," *V* for "vocabulary."

[3] One possible procedure is described in detail in A. O. Booth, L. Brandwood, and J. P. Cleave, *Mechanical Resolution of Linguistic Problems*, London, 1958, pp. 35–43.

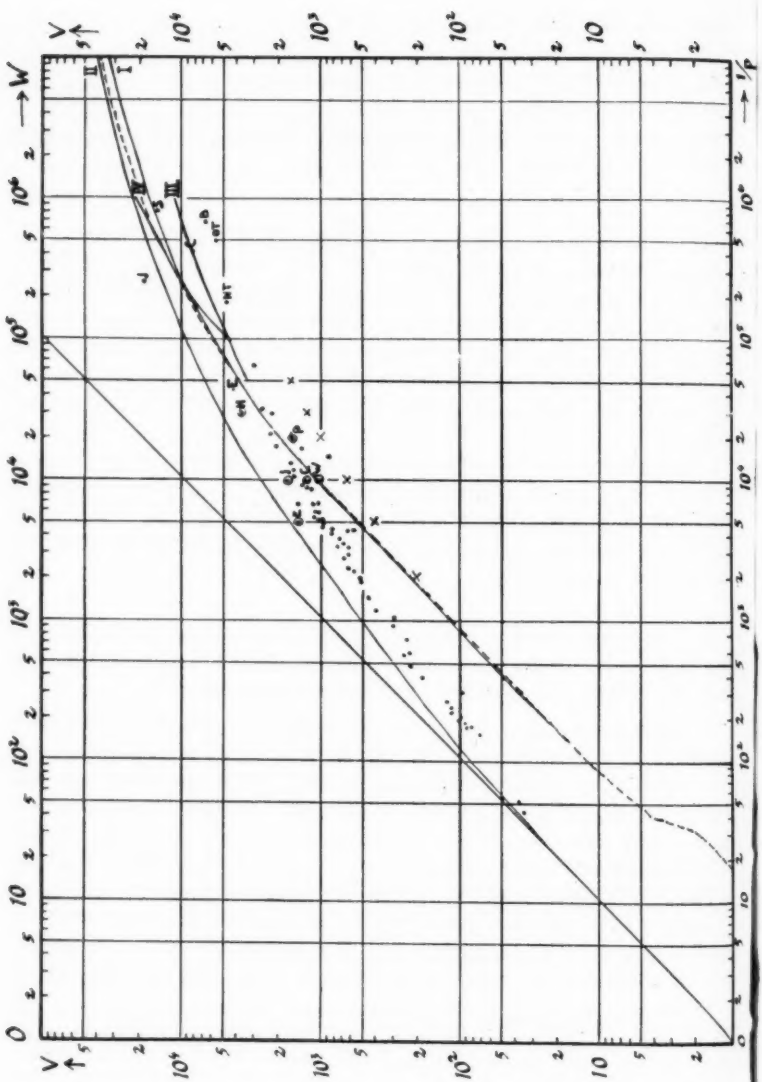


DIAGRAM 1. DISTRIBUTIONS OF LEXICAL UNITS.

fortunate, therefore, that it is possible to solve the problem, at least approximately, by a short-cut.

The solution consists in using the information provided by the compilers of frequency word lists. For English, the most complete of these lists is that of Thorndike and Lorge[4] (TL for short). It is based on a source material of about 18 million running words, combining four different counts covering four and one-half million words each. The first of these counts consisted of miscellaneous material, the second of American magazines, the third of juvenile books and school books, and the fourth again of miscellaneous material.

In the TL list the frequency of occurrence of each particular word is expressed in terms of the number of occurrences in a text mass of one million words. Needless to say, such a figure can be an average only. Whether we are to find, say, the word *chancery* in a text mass of one million words of course depends on the sort of books we include. (Actually, the word occurred, respectively, 12, 1, 1, and 11 times in the four $4\frac{1}{2}$ million counts making up the list). Therefore, when we say that the frequency of a particular word is 1 per million, we mean that in a text mass of 1 million words, *built up in precisely the same manner as the whole TL material*, with the same proportion of literary, general, and technical texts, the word in question is likely to appear once.

But this also implies that it will occur twice in a text mass of 2 million words, and in one only of two text masses of $\frac{1}{2}$ million each. More generally, in a text built up as a scale model of the TL material, the *probability of occurrence* of the word in question is 1/1,000,000, or 10^{-6}. We shall call this figure the word's *relative frequency*, and symbolize it p.

The *broken* line of Diagram 1 is a graphical representation of this aspect

---

Broken line: rank (V)-frequency (p) distribution of Thorndike–Lorge words (including proper names).

Curve I: rank-frequency distribution of lexical units in English, based on TL material.

Curve II: number of lexical units (V) per text mass (W), theoretical values based on TL material.

Curve III: number of lexical units per text mass, hypothetical average for homogeneous course.

Curve IV: number of lexical units per text mass, hypothetical average for heterogeneous course.

Dots: lexical units per text mass, Chaucer values.

Initialled dots and crosses: lexical units per text mass, other consecutive text values (see text, section 2, including note 18).

[4] E. L. Thorndike and I. Lorge, *The Teacher's Word Book of 30,000 Words*, New York, 1944.

of the TL list. The diagram is drawn on log-log paper, the abscissa indicating the relative frequency (actually, its inversion, $1/p$), and the ordinate indicating the number of words with such a relative frequency *or a higher one*. In order to find out how many words there are in a particular frequency range, for instance, the number occurring once per million (i.e., with p values between 1×10^{-6} and 2×10^{-6}) we subtract the V value corresponding to the higher p value from the one corresponding to the lower p value. In our example, this yields $19{,}440 - 14{,}231 = 5{,}209$.

The TL list counts proper names as words. In many ways this is reasonable and justifiable. Names like France, Waterloo, Socrates, and Hitler are meaningful units in any text, just as Jack and Jill are meaningful in the particular context where they occur. For several reasons, however, it is desirable to account for the proper names separately: one of them being that many vocabulary counts do not include them. I have therefore drawn Curve I in Diagram 1 in an attempt to represent the frequency distribution of the TL vocabulary when purged of proper names.[5] It is this distribution of *lexical units* which we shall be mainly dealing with from now on. The difference between the two curves is not very striking.

Our next question is: How many different lexical units will a text sample of one million words contain? It may be easy to jump to the conclusion that it ought to contain 17,000, since that is the number of lexical units likely to occur at least once in a one million-word text mass, according to Diagram 1. But such a conclusion would be wrong. For our assump-

[5] The curve is drawn on the basis of the following data. Three of the first 500 entries (0.6%), twelve of the following 500 (2.4%), and about 11.5% of the first 19,440 words were found to be proper names, leaving roughly 17,200 other lexical units above the frequency one per million. As will be found below, the first 500 entries contribute about 70% of the occurrences, the next 500 about 10%, and the remainder about 20% of the running words of a text sample. Hence the proper names will contribute about $0.006 \times 70 \times 0.024 \times 10 \times 0.115 \times 20 = 2.96\%$ of the occurrences due to the words with a frequency above one per million. Accordingly the removal of proper names lowers V by about 11.5% in the upper (right-hand) portion of the diagram, while W is reduced by about 3% only.

Professor Tord Ganelius, Göteborg, informs me that the curve representing the TL material including proper names can be described with remarkable accuracy by means of the simple formula:

$$e^{\frac{V}{7{,}520}} - 1 = \frac{W}{75{,}200}$$

which implies that for low values of W, V is one-tenth of W, and for high values of W, increasing W by a factor of e increases V by 7,520 units. (My own extrapolation of Curve I in Diagram 1, which excludes proper names, assumes that for high values of W, an increase by a factor of e increases V by 8,690 units. Curve I cannot be described in the same simple manner anyway.)

tion is that the one million word sample should be built up in the same manner as the total TL material. That is to say, the proportion of frequent to infrequent words in the sample should be similar. For instance, if the 100 most frequent words contribute 52% of the word occurrences in the whole material of 18 million words (as in fact they do), then they should also contribute 52% of the occurrences in a sample of one million, subject, of course, to random fluctuations.

Likewise for the infrequent words. The words used less than once per million in the TL material are indeed not frequent enough *individually* to appear in a scale model of the total material, but *as a group* they are.

As will be explained presently, we shall find that the words (lexical units) having a relative frequency of less than one per million contribute, as a group, 0.69% of the *running* words in any text mass built up in the manner of the TL material. The question, therefore, is this: how many *different* words will that be in a text sample of given size?

A fair approximation[6] is obtained by dividing the vocabulary into several portions, each consisting of words within a definite range of relative frequency, and then calculating the number of words contributed by each such portion by means of the Poisson formula:

$$V = L(1 - e^{-x})$$

where V is the number sought, L the total number of different words within the frequency range in question, e the base of the natural system of logarithms (2.71828 . . .), and

$$x = \frac{NP}{L}$$

where N is the size of the sample (number of word occurrences), and P the proportion of occurrences contributed by the frequency range in question.[7]

[6] The following simplified argument yields a rough estimate. In a text mass of one million words, all the 17,000 lexical units with relative frequency of 10^{-6} *or more* should occur at least once, and the words with a relative frequency below 10^{-6} should occur at most once (i.e., either once, or not at all. Hence all the occurrences of the rare words will be once-occurrences; and so the rare words, which contribute 0.69% of the running words (according to Table 1), should contribute 6,900 different words. This would bring the total to be expected up to $17{,}000 \times 6{,}900 = 23{,}900$. In fact, however, this reasoning yields too high a value for the number of words to be found in a given text mass. In a random sample of the TL material, some of the words with relative frequency above 10^{-6} would fail to appear, and some of the words with relative frequency below 10^{-6} would occur more than once. This is clearly brought out in our Table 2.

[7] I have taken over this formula from Guiraud, *Caractères Statistiques*, p. 33, to whom I refer for a fuller discussion of it. The symbol L is used by Guiraud for *Lexique*, or lexicon, the hypothetical potential vocabulary of which the actual vocabulary is but a

Ideally, we should have a great many such ranges, but for our present purposes four will be enough. The first range should consist of frequent words, which will all appear in the sample (i.e., $e^{-x}=0$). The second range should consist of words with a relative frequency which gives them roughly a fifty-fifty chance of occurring in the sample. The third range should consist of less frequent words, and the fourth of words with such a low frequency that their probability of occurrence in the sample is practically nil (i.e. $e^{-x}=1$). By this arrangement, only the two middle ranges need be actually calculated for each sample.

Our first step is therefore to estimate how big a proportion of the *running words* in a sample are contributed by each of these four frequency ranges. We do this from Curve I of Diagram 1, and draw up Table 1 which summarizes our findings.[8]

TABLE 1. FREQUENCY DISTRIBUTION OF ENGLISH LEXICAL UNITS.

Range down to p	V in range	average p in range	Σp in range	Σp cumulative (adjusted)	V cumulative
10^{-2}	12	1.33×10^{-2}	.2650	.2650	12
10^{-3}	109	2.67×10^{-3}	.2911	.5560	121
10^{-4}	969	2.58×10^{-4}	.2502	.8060	1,090
10^{-5}	4,810	3.08×10^{-5}	.1484	.9540	5,900
10^{-6}	11,100	3.47×10^{-6}	.03872	.9930	17,000
10^{-7}	16,000	3.71×10^{-7}	.00595	.9990	33,000
10^{-8}	20,000	3.95×10^{-8}	.00079	.9999	53,000
10^{-9}	20,000	4.50×10^{-9}	.00009	.99999	73,000
10^{-10}	20,000	4.50×10^{-10}	.000009	.999999	93,000

reflex (see below, note 20). It must be understood that the formula yields an approximation only, since it assumes that the probability of occurrence is the same for all the words belonging to the selected frequency range, which is not strictly the case.

[8] Clearly that proportion is simply the sum of the relative frequencies of each word in the range. If we assign rank numbers to all the words in the language, and let a be the rank of the most frequent word, and b the rank of the least frequent word in the range we are considering, we have

$$P = \sum_{V=a}^{V=b} p$$

When estimating P from the diagram, Curve I has been divided into small portions, for which a reasonably accurate average frequency (P) can be obtained by simply dividing the logarithmic distance between the two limits of the portion into two equal parts, and reading off the frequency corresponding to the point so obtained. Table 1 has been condensed for reasons of space: for the purposes of calculation, the curve portion

We next draw up Table 2, using the Poisson formula just referred to. The four frequency ranges[9] are, for each sample size N:

range 1 (first row): frequency down to $p = 1.43/N$
range 2 (second row): frequency down to $p = 6.7/10\ N$
range 3 (third row): frequency down to $p = 1/10\ N$
range 4 (fourth row): frequency below $p = 1/10\ N$.

TABLE 2. NUMBER OF LEXICAL UNITS (V) IN TEXT SAMPLES OF DIFFERENT SIZE.

Sample size N	Σp in range	L in range	$x = \frac{NP}{L}$	e^{-x}	$1-e^{-x}$	$L(1-e^{-x})$	V in sample
10^3	.5139	86				86	
	.0903	94	9.6×10^{-1}	.3829	.6171	57	
	.2019	910	2.22×10^{-1}	.8009	.1991	181	
	.1939	16,000	1.21×10^{-2}	.9880	.0120	192	516
10^4	.7673	770				770	
	.0716	725	9.89×10^{-1}	.3719	.6281	455	
	.1152	4,405	2.61×10^{-1}	.7703	.2297	1,010	
	.0459	27,000	1.7×10^{-2}	.9831	.0169	455	2,690
10^5	.9424	4,900				4,900	
	.0227	2,300	9.89×10^{-1}	.3719	.6281	1,440	
	.0280	9,800	2.86×10^{-1}	.7513	.2487	2,430	
	.0069					690	9,460
10^6	.99027	14,600				14,600	
	.00493	4,900	1.04	.3535	.6465	3,160	
	.00390	13,500	2.9×10^{-1}	.7483	.2517	3,400	
	.00090					900	22,060
10^7	.998852	30,800				30,800	
	.000551	5,700	9.67×10^{-1}	.3802	.6198	3,520	
	.000507	16,500	3.06×10^{-1}	.7364	.2636	4,450	
	.000090					900	39,670
10^8	.999865	50,000				50,000	
	.000036	3,000	1.2	.3012	.6988	2,100	
	.000090	20,000	4.5×10^{-1}	.6376	.3624	7,250	
	.000009					900	60,250
10^9							80,250

between $p = 10^{-2}$ and $p = 10^{-3}$ was divided into seven parts, whose limits were (p expressed in thousandths): 6.7, 5, 3.3, 2.5, 2, 1.43, and with the following averages: 8.2, 5.8, 4.1, 2.9, 2.2, 1.7, 1.2. Similarly for the lower frequency ranges. Later calculations in this paper have of course been based on this complete set of figures.

[9] See note 8, above.

The values of e^{-x} can be read off directly from the tables.[10] Note that the value of L in range 4 does not influence V. In this range all the words occur only once each.

Curve II of Diagram 1 has been drawn on the basis of the values in Table 2. The curve indicates the number of lexical units to be expected in a text mass of any given size, built up in the same manner as the total TL material. More exactly, it gives the most probable V values for random samples of that material. Ideally, such a sample is obtained by writing each running word of the total material on 18 million separate slips of paper, mixing the slips thoroughly in a bag, and, finally, drawing out a given sample at random from the bag.

2. We shall now see how actual *consecutive* texts compare with such theoretical mixed samples. For this purpose we shall use the data of published vocabulary counts. The comparisons, however, have to be handled with care, owing to the fact that different compilers have different concepts of what a word is.

The TL list on the whole followed ordinary dictionary practice, except for its inclusion of proper names. It is for this reason that we may regard Curve I of Diagram 1 as representing the distribution of "lexical units" in English, "lexical unit" meaning the same as "dictionary word"—which is itself, of course, a rather vague concept.

In general, when there is no information to the contrary, we may take it that writers on these matters understand "different words" as "lexical units." The most frequent alternative is to count each inflected form of a word as a separate word. We shall term these entities "word-form units." When dealing with such counts, we shall discuss how the number of lexical units can be estimated from the number of word-form units.

My first group of vocabulary values for consecutive texts is taken from Mersand's investigation of Chaucer's vocabulary.[11] Mersand calculated V and W (apparently in terms of lexical units) in each individual Chaucerian poem. I have indicated the values of each of the 59 poems by a dot in Diagram 1. Thus the dot at $V=3{,}032$: $W=64{,}908$ stands for *Troilus and Criseyde*, and the aberrant one at $V=884$: $W=14{,}779$ for the treatise on the *Astrolabe*. Finally, the dot at $V=8{,}430$: $W=420{,}000$, marked "C", stands for the whole of the Chaucer material in Mersand's count, which yielded a total Chaucerian vocabulary of 8,430 different words.

We note that the positions indicated for the individual texts form a

[10] E.g. *Tables of the Exponential Function* e^x, U.S. National Bureau of Standards, Applied Mathematics Series 14, Washington, 1951.

[11] J. Mersand, *Chaucer's Romance Vocabulary*, New York, 1939.

remarkably continuous series. Extrapolating the series towards the left it would meet the theoretical curve somewhere around $W=10$, $V=10$; i.e. at the point where that curve begins to coincide with the line $V=W$. That is quite in order: an actual text of connected speech containing no more than 10 words—an ordinary short sentence, for instance—will normally consist wholly, or almost wholly, of different words. The reader may test this for himself by counting, on this page, the number of words which occur between repetitions of any one word.

But the greater the text size, the farther do the Chaucer values fall short of the theoretical ones. On an average, a 1000 word Chaucer text reaches 60 to 65% of the theoretical V value, a 10,000 word text only 50 to 55%, and the 64,000 word *Troilus and Criseyde* only about 40%. The total Chaucer corpus attains about 50% of the theoretical V value.

The reason for the smaller vocabularies in the consecutive texts as compared with the theoretical samples is pretty obvious. A consecutive text deals with a limited subject-matter: the writer, accordingly, has no reason to range over the very wide and disparate area of knowledge covered by the TL material.

If we interpolate an average V,W curve for the *individual* Chaucer text, and then extrapolate towards the right, we shall find that the V value for the complete works falls above the extrapolated curve. This is as it should be: a collection of different works, even by the same writer, is likely to cover a wider field than any single consecutive work of comparable length.

Somebody may object that it is misleading to compare Chaucer's texts with the theoretical distribution of random texts drawn from (mainly) modern material. However, though the Middle English vocabulary was different from the modern one, there is every reason to suppose that it was distributed in roughly the same manner. A theoretical curve based on Middle English material would no doubt follow the present curve very closely, at least in the portions below $W=10^5$. In fact, as we shall see below, the Chaucer values might almost serve as standard or normal for average consecutive modern texts.

Our next illustrations concern Shakespeare and the English Bible. One estimate (uncertain) of Shakespeare's vocabulary puts it at 15,000 words.[12] Assuming that this estimate is based on what we call lexical units, and estimating the Shakespeare corpus at about 800,000 running words, Shakespeare's vocabulary would be about 72% of the theoretical value, considerably higher than Chaucer's. I have indicated the position by "S" in Diagram 1.

[12] O. Jespersen, *Growth and Structure of the English Language*, Oxford, 1952, pp. 199–200.

Counts of the Authorized Version of the English Bible have yielded 5,642 different words in the Old Testament, 4,800 in the New, and 6,568 in the whole Bible.[13] Estimating the Old Testament at 500,000 running words, and the New at 180,000, we get the positions indicated by "OT", "NT", and "B" in Diagram 1; respectively 32%, 40%, and 33% of the theoretical values.

We next come to those vocabulary counts where the unit counted is the word form, not the lexical unit. The fullest of these is Hanley's Word Index to James Joyce's *Ulysses*, which yielded 29,899 word-form units in a text mass of 260,430 running words.[14] A sample count of 20 columns in the list showed that the lexical units formed 61.5% of the entries (standard error $\pm 1.9\%$). Hence the V value in terms of lexical units is 18,400. As to the W value, a sample count indicated that about 13.4% of the list entries were proper names. But as the names are by and large less frequent than other words, the W value should be reduced by much less than 13.4%. Using a similar argument to the one employed above,[15] we arrive at a reduction value of 4.45%. Hence James Joyce's position, indicated by "J" in Diagram 1, is $V = 18{,}400$, $W = 246{,}000$, which places him about 32% *above* the theoretical V value for a theoretical sample text of that size.

The figure we have used for reducing the V value of the Hanley list to lexical units cannot be accepted as generally valid. One difference is that *Ulysses* has an unusually high number of proper names: hence the reduction figure of 61.5% is unusually low. For comparison we may consider a count of the English Bible, which yielded 6,568 lexical units, and 9,884 word-form units.[16] That yields a percentage of 67.5%.

There is one more complication to be taken into account, namely, the length of the text. For very small texts, V counted in terms of word-form units is identical with V counted in lexical units, for the simple reason that most words have time to occur in one form only. The discrepancy between the two ways of counting grows gradually larger with increasing size—as the proportion of once-occurrences in the vocabulary of the text grows gradually smaller.[17]

[13] *Ibid.*

[14] Miles L. Hanley, *Word Index to James Joyce's Ulysses*, Madison, 1951.

[15] See above, note 5.

[16] O. Jespersen, *Growth and Structure*, p. 200. I assume that the count excluded proper names.

[17] For the counts presented below, which all include proper names, I employ the following figures, arrived at after some trial and error: $W = 5{,}000$, reduction percentage 90%; $W = 10{,}000$, 85%; $W = 20{,}000$, 80%; $W = 30{,}000$, 75%; $W = 50{,}000$, 70%; $W =$ 100,000 or more, 65%.

The values for the remaining texts included in Diagram 1 have been estimated in accordance with the above argument.[18]

3. It is now time to discuss the relation between the theoretical V values, based on thoroughly mixed samples, and the observed values, based on consecutive texts.

In the first place we note that, except in one case, the V values for consecutive texts are smaller, and often much smaller, than the theoretical values, which confirms the argument we advanced above, in our discussion of the Chaucer material. The exception, in this case, may with entire propriety be said to prove the rule. James Joyce's *Ulysses* is beyond question a highly unusual literary product. It would have been rather disappointing if our test had not showed it up as such.[19]

As regards the other works listed we notice that T. S. Eliot reaches about 85% of the theoretical value, and Hamlet about 72%, almost exactly the same as our figure for the entire Shakespearian vocabulary. Johnson, Congreve, and the Paston letters fall roughly in the Chaucer range, Wycherly somewhat below, while the schizophrenic patient exhibits an exceedingly poor vocabulary: around 25% of the theoretical value.

The newspaper sample reaches about 65%. That is a fairly high figure, the result of two factors working in opposite directions: the limited

[18] The following data are from G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Cambridge, Mass., 1949), pp. 25, 123, 291. One set of data concerns 44,000 words of American newspaper material, carried out by Eldridge. It yielded 6,002 word-form units, which I equate with 4,200 lexical units. Reducing the W value by 3% for proper names, we arrive at the position indicated by the dot at "E" in Diagram 1.

The ringed dots in Diagram 1 represent the following texts: Paston Letters ($W=20{,}000$, $V=1{,}600$ lexical units); Hamlet ($W=30{,}000$, $V=3{,}750$); Wycherley ($W=10{,}000$, $V=1{,}020$); Congreve ($W=10{,}000$, $V=1{,}280$); Johnson, Rasselas ($W=10{,}000$, $V=1{,}700$); T. S. Eliot ($W=5{,}000$, $V=1{,}440$).

The crosses in Diagram 1 represent six different samples of letters written by a schizophrenic patient—I: $W=50{,}000$, $V=1{,}640$; II: $W=30{,}000$, $V=1{,}270$; III: $W=20{,}000$, $V=1{,}000$; IV: $W=10{,}000$, $V=650$; V: $W=5{,}000$, $V=405$; VI: $W=2{,}000$, $V=210$.

[19] Zipf, op. cit., p. 546, held that the "Ulysses data provided the most clear-cut example of a harmonic distribution in speech of which I know," and regarded them as confirmation that his "law," namely, rank × frequency = constant (the constant being 1/10) was valid for large samples as well as for smaller ones. As a matter of fact, *Ulysses* is too exceptional a text to serve as a standard of reference. The other text data given above show that Zipf's law holds only for a limited range of text sizes. Expressed in our terms, Zipf's law states that $Vp=0.1$. This implies that the values for individual texts should all fall on a straight line with slope 45°, starting from $V=1$, $p=10^{-1}$. That is approximately true of samples between $W=5{,}000$ and $W=50{,}000$ (or rather more if Zipf's word definition is used). But above that region the value of Vp decreases more and more: for $W=10^6$ it is 0.022, with our word definition. For a discussion of Zipf's law, see also Guiraud, *Caractères Statistiques*, pp. 12, 24.

vocabulary to be expected in a newspaper is counteracted by the diversity of the material, which does not consist of one consecutive text, but of several smaller and unconnected pieces.

We now possess the requisite data to consider more fully the problem of how to estimate the vocabulary of a given text mass. The theoretical curve may be taken to indicate the number of words to be expected in text samples of maximum diversity, consisting of material from all possible kinds of literature.

Only very unusual consecutive texts will have as large vocabularies as the corresponding mixed sample. The fact that James Joyce's *Ulysses* actually exceeds the theoretical value is a corollary of Joyce's stylistic technique, which, by means of associative linkings, attempted to bring the whole complexity of one man's experience into the space of a book treating of a single day's happenings. The *Ulysses* vocabulary warns us that the theoretical values are not theoretical maxima. The real maximum values would of course be represented by the straight line standing for the function $V = W$. Such values would be reached by some very special texts, namely, dictionaries.

Most ordinary consecutive texts will contain much smaller vocabularies than the theoretical ones. The range of variation is very great, but for *individual writers* it seems fairly narrow. At least for text masses between about 20,000 and one million running words, the observed vocabularies remain fairly constant as measured in per cent of the theoretical values. Hence these percentages may to some extent be used as a measure of the richness, or variety, of an author's vocabulary, independently of the text size.

In these terms, Shakespeare's vocabulary is more varied than Chaucer's. It is characteristic of Shakespeare to bring into play a wider range of words than Chaucer does. As for the total vocabulary *known* by either author, we of course know nothing directly. The vocabulary actually employed depends very largely on the subject matter. An author who concentrates on a limited range of subjects has less occasion to employ a large vocabulary than one who continually ventures into new fields. On the other hand, it is fair to add that a writer's choice of subject matter is to some extent a function of his range of experience, and so not unrelated to his personality traits. Moreover, a man with a wide experience is likely to bring his experience to bear on his treatment of any subject. For this reason the vocabulary of an author, expressed in the way we have proposed, will normally be fairly representative not only of the actual, but also of the potential vocabulary.[20]

[20] Another way of estimating a writer's vocabulary has been suggested by Guiraud. Guiraud makes a distinction between the *vocabulary*, *V*, the words appearing in the

Needless to say, the vocabulary used by an author depends also to a great extent on the resources of the language he employs. Hence Chaucer's smaller vocabulary as compared with Shakespeare's may to some extent be due to the more limited resources of Middle English as compared with Elizabethan English. The vocabulary measure we have proposed measures directly only the relative richness of the *actual* vocabulary of the text. In order to interpret the figure, many additional factors have to be taken into account.

Though our material is scanty, it seems a fair guess that the works of an "average" English writer will contain vocabularies corresponding to about 50% of the values obtained from a thoroughly mixed sample—somewhat higher for small texts, and somewhat lower for large ones. Hence we should expect values falling around Curve III in Diagram 1, which represents V values 50% below Curve II.

4. We now turn to the following question. Consider the case of a foreign language student who is given the task of reading through 10 volumes of 100,000 words each. Supposing he learns and remembers all the words in that material, how large will his vocabulary be after the first, second . . . etc. book?

If the books are ordinary ones, and all by the same author, we may take Curve III to give the answer. The same is true if the books are by different authors, but deal with similar subjects in a similar manner. But if each book

---

actual text, and the *lexicon*, L, the hypothetical potential vocabulary on which the writer draws when composing his text—the "words at risk," as Yule said. Guiraud estimates L from the ratio of once-occurrences to the total number of words in the vocabulary. Further, he estimates the *concentration*, C, of the writer's vocabulary by dividing the number of occurrences contributed by the 50 most common words, by the total number of words in the sample. Finally, he measures the *dispersion* or *richness*, R, of the vocabulary in terms of the formula $V/\sqrt{N}$. Both R and C are in their turn related to standard values and to L.

Guiraud establishes beyond doubt that L, R, and C, and their derivatives, are (approximately) constants for individual authors, at least for text sizes around 5,000–100,000 words. My own material confirms this in the case of R: the Chaucer value is around 14 within that range. I am somewhat more doubtful about the interpretation of those of Guiraud's constants which involve L. L itself is calculated on the assumption that the writer's lexicon, whether great or small, has the same *proportion* of words within each frequency range as the "normal" lexicon of 24,000 words. That seems hardly realistic; it appears more likely that writers with a large vocabulary differ from those with a smaller one mainly in the number of words belonging to the infrequent ranges. However, as Guiraud himself stresses, whatever the interpretation of L, there is no doubt that it is a characteristic constant of a writer's language, and any discussion of these matters must take account of Guiraud's important and stimulating contribution.

is widely different from every other, both in style and in subject matter, we must expect the student's vocabulary to increase much more rapidly as he passes from one book to another. Indeed, it is possible that ten ordinary books, suitably selected, might more or less equal the vocabulary of a thoroughly mixed sample of the same size. Curve IV of Diagram 1 represents this hypothesis. Table 3 gives the actual figures for the two alternative courses, as read from Curves III and IV.

TABLE 3. PROBABLE VOCABULARIES OF LEXICAL UNITS IN TWO DIFFERENT SETS OF TEN ORDINARY ENGLISH BOOKS.

	Homogeneous course		Heterogeneous course	
Book no.	Cumulative total	New words	Cumulative total	New words
1	4,750	4,750	4,750	4,750
2	6,300	1,550	8,600	3,850
3	7,400	1,100	11,000	2,400
4	8,250	850	13,200	2,200
5	9,000	750	15,000	1,800
6	9,600	600	16,500	1,500
7	10,100	500	17,900	1,400
8	10,600	500	19,200	1,300
9	11,000	400	20,200	1,000
10	11,300	300	21,000	800

Let us summarize the simplifying assumptions on which the estimates are based:

1. The vocabulary of each individual book is of normal size, with the normal proportion of frequent to infrequent words. Each book, therefore, is assumed to have a vocabulary of 4,750 lexical units.

2. In the homogeneous course, the books do not differ more from each other than two books by the same author normally do.

3. In the heterogeneous course, the books differ from each other as much as possible, within the limits of the TL material.

Clearly both courses represent extremes; a random sample of ten books would naturally be likely to yield values somewhere between the two. It is for this reason that I have thought it worth while to work out the argument in detail, in spite of the large amount of guesswork involved.

5. Up to now, our discussion has been based on the TL word definition, modified to cover what we have called the lexical unit. For many purposes,

however, the lexical unit is an unsuitable basis of operations. That is the case, for instance, with the problem which we have just touched upon, namely, the process of acquiring a vocabulary in a foreign language.

The trouble is that no difference is made between root words and derivations, while it is obvious that many derivatives will be "known" by the learner even if he has never met with them in his previous reading. In other words, some derivations, at least, are on a par with regular inflections from the semantic point of view. Therefore, if our aim is to predict the extent to which the meaning of the words in a random text will be known, we should base our count, not on words as usually defined, but on "semantic elements." Unfortunately such a count would involve pretty formidable difficulties. What are semantic elements? The linguist's morphemes are likely candidates: but their definition is also controversial. In particular, it is hardly likely that a purely formal definition of the morpheme would consistently single out the semantic elements we are after. What about *-t* in *height* as against *high*, as compared with *-t* in *sight* as against *see*? What about the meaning of *ob-* in *obtain*, *oppose*, *obviate*, *obstacle*, etc.?

It seems clear that if minimum semantic elements are what we seek, they will not coincide exactly with either etymological or formal morphemes. Moreover, it is fairly certain that what appears as a semantic element to one speaker may not be so to another: it depends on his linguistic imagination, and on his knowledge of the language. In general, the better a person knows a language, the better will he be able to analyse its words into semantic elements, and thus to guess at the meaning of new words formed after a definite pattern.

A further complication is the fact that vocabulary counts normally do not distinguish between words of one meaning and words with several meanings. Yet from the point of view of the language learner, a form with two distinct meanings will have to count as two semantic elements. But what are two distinct meanings? We are not interested in such semantic variations as can be readily inferred from the context. We wish to reckon two meanings as distinct only if one cannot be inferred from the other by means of the context. On this point, again, there are bound to be individual variations. What one person regards as distinct meanings, the other will regard as nothing but contextually conditioned semantic variations. What are, for instance, the semantically independent entities conveyed by the word *nice*?

Yet though the establishment of a vocabulary of semantic elements will certainly meet with formidable difficulties, I do not think they are so great as altogether to rule out the approach. Though there are bound to be individual differences, it ought to be possible to establish the average

"semantic ability," or if we wish, a whole series of averages valid for different groups of speakers with different degrees of linguistic proficiency. A questionnaire technique might be employed for the purpose.

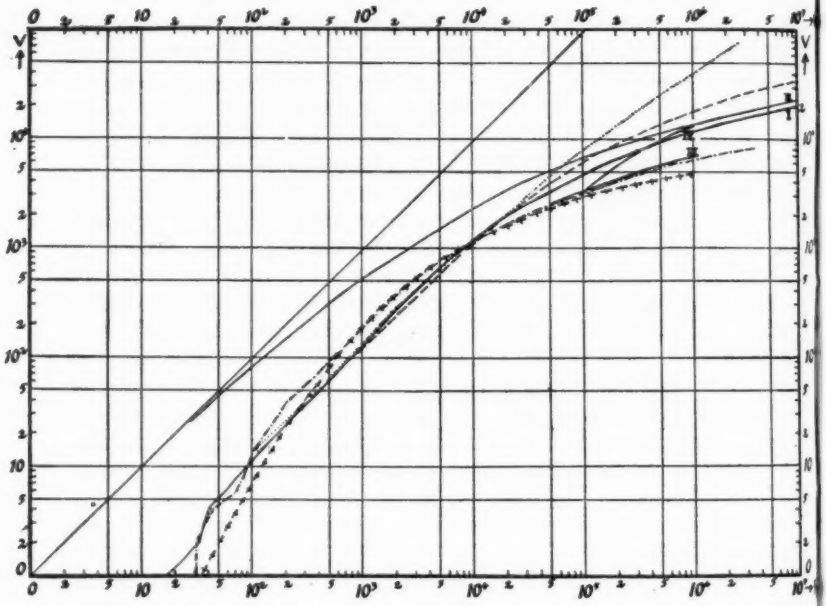


DIAGRAM 2. DISTRIBUTIONS OF SEMANTIC ELEMENTS.

Broken line: rank (V)-frequency(p) distribution of lexical units in English, based on TL material.

Curve I: rank-frequency distribution of semantic elements in English, based on TL material (see text).

Curve II: number semantic elements (V) per text mass (W), theoretical values based on TL material.

Curve III: number semantic elements per text mass, hypothetical average for homogeneous course.

Curve IV: number semantic elements per text mass, hypothetical average for heterogeneous course.

Dotted line: rank-frequency distribution of German word-form units (Kaeding's data).

Stroke-dotted line: rank-frequency distribution of German semantic elements (Kaeding's stem syllables).

Cross line: rank-frequency distribution of Chinese characters (Zipf's data).

As there does not as yet exist any body of experimental knowledge on these matters, I have myself examined the TL list from this point of view. A sample count indicated that about 60% of the words (excluding proper names) with $p > 10^{-6}$ were such as could not be understood from a knowledge of more common related forms. The corresponding figure for the less frequent words in the TL list turned out to be 45%, while for the 1,000 most frequent words it is as high as 96%.[21]

Starting from this information, we now attempt to reconstruct the frequency distribution of the semantic elements in the language, on the basis of the TL list. There is a complication here. The removal of, say, *probably* must be accompanied by the addition of its relative frequency to that of the word *probable*. Generally, the relative frequency of the more frequent words will increase, and the aggregate frequency of the infrequent ones will decrease. The sum of the relative frequencies of all semantic elements, like that of the lexical units, must of course be 1.[22]

On this basis I have drawn up Curve I in Diagram 2. Curve I of Diagram 1 is reproduced in Diagram 2 as a broken line, for comparison's sake.

6. Partly to check the results of the above argument, and partly because of their intrinsic interest, I have at this point brought in data from other languages for comparison.

A German frequency count by Kaeding is especially relevant in this connection.[23] Like Zipf and Hanley, Kaeding counts word-form units, but he provides in addition frequency lists of word stems, prefixes, and endings. His material consists of a text mass of 10,900,000 words. Personal names, but not place names, are included.

The dotted line of Diagram 2 is constructed on the basis of Kaeding's data in exactly the same way as the broken line of Diagram 1 from the TL data: it sets out the number of word-form units V with a given relative

[21] Polysemy had to be left out of account. It should be noted in this connection that it is a problem that concerns above all the most frequent words in the language. The less frequent a word is, the less likely is it to have several meanings (Zipf, *Human Behavior*, pp. 27ff. Guiraud, *Caractères Statistiques*, p. 2).

[22] In this calculation we assume that one or more lexical units always yield one, and only one, semantic element. We do not list endings, prefixes, etc., as semantic elements. This is necessary because the frequencies are expressed in terms of the number of occurrences of word units. To convert this into number of occurrences of semantic elements proper appeared too baffling a task. The procedure adopted also implies that compounds can be referred to one of the components only in the count of semantic elements.

[23] F. W. Kaeding, *Häufigkeitswörterbuch der deutschen Sprache*, Steglitz bei Berlin, 1898.

GEN. LIBRARY UNIV. OF N. DAK.

frequency p or a higher one.[24] In the same manner, the stroke-dotted line sets out the frequency distribution of the German word stems, taken direct from Kaeding.[25]

The family likeness between the English and German distributions is evident. They both have a slope of about 45° below $V = 1{,}000$, after which the slope gradually decreases. Further—and this is the main point of the comparison—the relation between word-form unit distribution and the stem distribution corresponds to the relation between lexical units and semantic elements, and thus confirms our argument above.

But the differences between the German and English distributions are equally marked and interesting. The German word distribution curve bends downwards much more slowly. That is not wholly due to the fact that it is based on word-form units, not on lexical units. A sample count of the words used four times or more in Kaeding's material showed that the word-form units could be reduced to about 41.3% as many lexical units. That yields 33,000 lexical units for the frequency above 3.7×10^{-7}. The corresponding Thorndike value is 24,000 lexical units. The reduction in percentage to 41.3%, as compared with the English one of about 65%, is a reflection of the inflectional nature of the German language as compared with English. The remarkably high number of different words in German (assuming it to be real: the reduction is not *strictly* correct) may be due to the facility of that language to form compounds. To take a simple example: while the German list will include as separate words *Wort*, *Buch*, *Häufigkeit*, *Wörterbuch*, *Häufigkeitswörterbuch*, the English will normally have just *word*, *list*, and *frequency*, but no compounds.

It is difficult to see why German should apparently work its 1,000 most common words more heavily than English does—for that is the implication of the German curve running above the English one in this portion. It may be due partly to the fact that the TL list includes common contractions, such as *I'll*, *can't*, etc., as separate words in addition to *I*, *will*, *can*, and *not*, whereas the German list only includes the full forms.

Kaeding's stem syllables correspond, by and large, to what we have called semantic elements. A sample count of the word list confirmed this: 9.4% of the 80,000 word-form units above $p = 3.7 \times 10^{-7}$ could be classed as semantically "independent." To the resulting figure, 7,500, should be

[24] Kaeding's list included personal names, but not place names. I have drawn the curve only to $1/p = 2.6 \times 10^6$, i.e. to include the words used four times in the material of 10.9×10^6 words.

[25] Kaeding's "Stammsilben" were not etymological stems, but, by and large, the stressed syllable of the word. Thus *hauf* (*as in Haufen*) was distinguished from *häuf* (*as in häufig*), and *gnos* was put down as the stem syllable of *Diagnose*.

added the 240 semantic elements yielded by the 320 most frequent word-form units (which were not included in the former list), whereby we arrive at about 7,800 semantic elements, which coincides almost exactly with the actual figure for stem syllables above that frequency.

We may probably rely on the stroke-dot curve in Diagram 2 as giving quite a good representation of the frequency distribution of semantic elements in German. Assuming this to be the case, we must consider its relation to the hypothetical English curve I of the same diagram.

The German values are only around 50% of the English ones for the lower frequencies. Hence the number of semantically independent words to be expected in large text masses of English will be about double that in German texts of similar size.

Is that a plausible result? It probably is. We have already seen that as regards the number of different lexical units, German takes the lead, largely because of its freedom of compounding and prefixing. But the word elements in German are usually perspicuous, and above all, the language makes do with its native stem material to a considerably greater extent than English does. To take an example: German has the grammatical term *Hauptwort*, and the element *Hauptworts-* corresponding to the English *noun*, *nominal*, and *substantive*, which are all semantically independent in our sense. This is the general rule: the English words are largely independent of each other, while the German ones exhibit their relationship to each other. German concentrates on the common Germanic material; English draws freely on Germanic, French, and Latin.

Further evidence to support our argument about the run of Curve I in Diagram 1 is provided by Zipf.[26] It concerns the frequency distribution of about 4,500 Chinese characters in a sample of nearly one million drawn from Chinese newspapers. As the characters largely function as semantic elements, it is hardly surprising to find that the curve (drawn as a line of crosses in Diagram 2) is very similar to the one valid for German stem syllables. Being based on a diagram in Zipf, the curve is approximate only.

7. After this somewhat roundabout testing,[27] we may accept our Curve I as a fair representation of the frequency distribution of English semantic

[26] Zipf, *Human Behavior*, p. 89.

[27] Hanley's *Ulysses* list provides a further check. My sample count of that list indicated that 38.9% (standard error ±2%) of the items could be reckoned as semantic units in our sense, yielding 11,600 units for the whole text of 246,000 words excluding names, as compared with 18,400 lexical units: a percentage of 63%. For a text of this size. Diagram 1 gives $V = 14{,}000$, and Diagram 2 gives $V = 9{,}000$, a percentage of 64%

TABLE 4. FREQUENCY DISTRIBUTION OF ENGLISH SEMANTIC ELEMENTS.

Range down to p	V in range	average p in range	$\sum P$ in range	$\sum p$ cumulative (adjusted)	V cumulative
10^{-2}	12	1.33×10^{-2}	.2647	.2647	12
10^{-3}	113	2.62×10^{-3}	.2960	.5607	125
10^{-4}	1,075	2.73×10^{-4}	.2922	.8529	1,200
10^{-5}	3,200	3.70×10^{-5}	.1181	.9710	4,400
10^{-6}	7,400	3.43×10^{-6}	.0253	.9963	11,800
10^{-7}	8,100	4.08×10^{-7}	.00330	.99960	19,900
10^{-8}	8,900	4.04×10^{-8}	.00036	.99996	28,800
10^{-9}	9,000	4.00×10^{-9}	.000036	.999996	37,800
10^{-10}	9,000	4.00×10^{-10}	.0000036	.9999996	46,800

TABLE 5. NUMBER OF SEMANTIC ELEMENTS (V) IN TEXT SAMPLES OF DIFFERENT SIZE.

Sample size N	$\sum p$ in range	L in range	$x = \frac{Np}{L}$	e^{-x}	$1-e^{-x}$	$L(1-e^{-x})$	V in sample
10^3	.5138	86				86	
	.1001	104	9.61×10^{-1}	.3825	.6175	62	
	.2390	1,010	2.39×10^{-1}	.7874	.2126	213	
	.1471					147	508
10^4	.8119	860				860	
	.0820	840	9.76×10^{-1}	.3768	.6232	523	
	.0771	2,700	2.86×10^{-1}	.7513	.2487	670	
	.0290					290	2,343
10^5	.96384	3,800				3,800	
	.01457	1,500	9.72×10^{-1}	.3783	.6217	933	
	.01793	6,500	2.76×10^{-1}	.7588	.2412	1,565	
	.00366					366	6,664
10^6	.99454	9,000				9,000	
	.00278	2,700	1.03	.3570	.6430	1,730	
	.00225	6,900	3.26×10^{-1}	.7218	.2782	1,920	
	.00043					430	13,080
10^7	.999439	18,800				18,800	
	.000288	3,000	9.61×10^{-1}	.3825	.6175	1,850	
	.000229	8,000	2.86×10^{-1}	.7513	.2487	1.985	
	.000044					440	23,075
10^8							32,075
10^9							41,075

elements. We next draw up Table 4 on the basis of the curve, and Table 5 on the basis of Table 4, in exactly the same way as we drew up Tables 1 and 2, above.

From Table 5 we get the values needed for Curve II of Diagram 2, which indicates the number of semantic units to be expected in a mixed TL sample of given size. Further, Curves III and IV of Diagram 2 are arrived at by the same argument as curves III and IV of Diagram 1, and serve as the basis of Table 6, which sets out, after the fashion of Table 3, the number of new semantic elements to be learnt from a thorough study of ten ordinary books, all books being similar in one course, and dissimilar in the other.

TABLE 6. PROBABLE VOCABULARIES OF SEMANTIC ELEMENTS IN TWO DIFFERENT SETS OF TEN ORDINARY ENGLISH BOOKS.

	Homogeneous course		Heterogeneous course	
Book no.	Cumulative total	New words	Cumulative total	New words
1	3,300	3,300	3,300	3,300
2	4,200	900	5,600	2,300
3	4,800	600	7,400	1,800
4	5,300	500	8,600	1,200
5	5,700	400	9,700	1,100
6	6,000	300	10,600	900
7	6,250	250	11,400	800
8	6,450	200	12,100	700
9	6,650	200	12,700	600
10	6,800	150	13,100	400

The values in the columns for "New words" in Table 6 are only about half as large as those in Table 3, the proportion growing even smaller as the number of books increases. This shows the influence of the fact which we drew attention to in our discussion above, namely, that the learner will understand many new words from his previous knowledge of cognates. A comparison of the tables indicates that while there are indeed likely to be about 1550 new words in the second book of the homogeneous course, no less than 1,550–900 = 650 of them will be readily interpreted by the reader on the basis of cognates. The simplifying assumptions underlying this argument were explained above, section 4, end.

8. We now possess some preliminary notions for estimating the vocabulary likely to be found in text masses of different sizes and different

composition, and also for comparing the vocabularies of individual texts with each other, independently of the text size.

Our next and last problem concerns the potential known (passive) vocabulary of an individual, and the methods that can be used to estimate it.

Previous vocabulary tests have on the whole led to the somewhat discouraging conclusion that the size of vocabulary of the person tested is largely determined by the size of the dictionary from which the test sample has been taken.[28] Usually the text has consisted of a random sample of words from a particular dictionary, and the percentage of words identified in the sample has been accepted, within the limits of sampling errors, as indicating the number of words known in the dictionary.

The argument is sound as far as it goes, but it does not go far enough. What is needed is a way of ascertaining how the vocabulary of the dictionary stands to that of the language as a whole. But if that vocabulary is infinite—and according to our argument above there is no need to assume any limit—the procedure seems indeed to lead to an *impasse*.

But there is a way out of the dilemma, and the way is indicated by our discussion on the preceding pages. Instead of trying to estimate an individual's vocabulary in relation to the language as a whole—which seems impossible—or in relation to a particular dictionary—which is rather futile—we may estimate it in relation to the number of words extant in the language *within a definite frequency range*. The number, and even more the actual words, composing the frequency ranges must be ascertained from frequency counts, based on very extensive and methodically selected material. For the present, however, we have to rest content with such lists as Thorndike–Lorge's.

The proportion of words a person is likely to identify in a sample list is evidently not the same in all the frequency ranges. Most persons will know all the words in the higher frequency ranges, and a gradually decreasing proportion in the lower ones. Ideally, therefore, we should try to ascertain what may be called the *vocabulary profile* of each language user, indicating the percentage of words known within a definite set of frequency ranges.

I have used this approach in order to assess the English vocabularies of Swedish students of English, basing the test as far as possible on what I have called semantic elements. The TL list was used in the following manner. First, of course, proper names were excluded. Second, such words as *emotion*, *emotional*, *emotionally* were counted as one unit, with a

[28] For a discussion of work on this subject, see Dorothea McCarthy "Language Development in Children," in *Manual of Child Psychology*, ed. Leonard Carmichael, New York, 1954, pp. 492–630.

frequency of $34+9+1=44$ per million (4.4×10^{-5}). In a group of semantically interdependent words, the most frequent one was chosen to represent the whole group (i.e. the semantic element): *emotion*, in the present example. Third, as the object was primarily to assess the effect of the students' reading of English, more or less international loanwords, such as *telephone*, *electricity*, *radio*, etc., were excluded. No attempt was made to estimate the number of such words, and their influence on the test results. Ten test words were selected at random from each of six frequency ranges, extending from $p=2\times10^{-4}$ to $p=2\times10^{-7}$.[29]

Knowledge of the word was tested by asking the students to give an English or Swedish synonym, or provide a context explaining the meaning of the word. The answers were graded +1 for correct, 0 for incorrect or no response, and ½ for doubtful cases. No attempt was made to grade the answers for fullness: if one correct meaning was given, it was considered enough.

The test evidently concerns passive knowledge only. But as such it is fairly exacting, since no context was provided at all. (Experiments confirmed that the provision of a context very markedly improved the test results.) As always when frequency counts are concerned, polysemy had to be left out of account. Several tests were constructed in the same way, using different words in each. The results showed a fair consistency, confirming the reliability of the test.

Table 7 sets out the test results, in terms of per cent words known in each frequency for groups of students with different standards of proficiency in English. Group I consisted of university students who had been doing English for at least 3 terms at the university (and sometimes been to

[29] The easiest procedure would be to use the frequency ranges described above, note 8. The average frequency in the range is obtained by dividing the aggregate proportion of occurrences ($\sum p$) in the range by the number of words in the range. For the range 10^{-5} to 10^{-6} this yields $0.02530/7{,}400=3.42\times10^{-6}$. Hence the test words for the range should be taken from those listed by Thorndike as occurring 3 or 4 times per million.

In actual fact, the test words were selected before the theory set out above was worked out, and the limits of the frequency ranges had to be chosen so as to fit the already selected averages, which explains the somewhat unsystematic appearance of the range divisions. The following are the actual words used in one of the tests: 1.22×10^{-4} (120–130 per million): *action*, *cause*, *contain*, *desire*, *finally*, *important*, *light*, *member*, *often*, *probably*. 5.6×10^{-5} (55–65 per million): *accident*, *affect*, *castle*, *choice*, *coast*, *defeat*, *introduce*, *meal*, *nail*, *occur*. 2.2×10^{-5} (20–30 per million): *witch*, *wreath*, *amendment*, *array*, *bewilder*, *bough*, *broad*, *canvas*, *challenge*, *chop*. 5.5×10^{-6} (5 per million): *arbitration*, *bail*, *cadence*, *deluge*, *ember*, *flop*, *gaudy*, *hag*, *lair*, *maim*. 1.3×10^{-6} (1 per million): *arable*, *banister*, *cad*, *dally*, *easel*, *feint*, *gamut*, *halibut*, *laity*, *manse*. 4.2×10^{-7} (less than 1 per million): *accretion*, *blarney*, *captions*, *daft*, *eavesdrop*, *fairway*, *flotsam*, *garble*, *halcyon*, *indict*.

TABLE 7. PER CENT SEMANTIC ELEMENTS KNOWN IN DIFFERENT FREQUENCY RANGES, IN SIX GROUPS OF STUDENTS WITH DIFFERENT STANDARDS OF PROFICIENCY

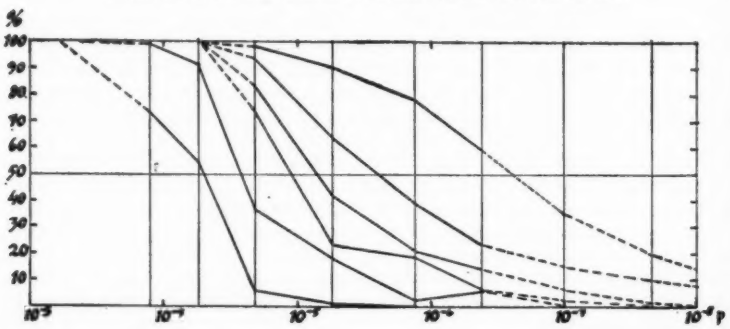
Group number	Group size	Group characteristics
I	6	70–90% in $p = 1.3 \times 10^{-6}$
II	22	30–50% in $p = 1.3 \times 10^{-6}$
III	18	30–50% in $p = 5.5 \times 10^{-6}$
IV	9	10–30% in $p = 5.5 \times 10^{-6}$
V	8	30–50% in $p = 2.2 \times 10^{-5}$
VI	12	40–60% in $p = 5.6 \times 10^{-5}$

	Per cent words known in p range					
	1.22×10^{-4}	5.6×10^{-5}	2.2×10^{-5}	5.5×10^{-6}	1.3×10^{-6}	4.2×10^{-7}
I	—	—	97.6	90.9	78.3	60
II	—	—	93.7	64.2	38.9	23.2
III	—	—	83.3	42	20.6	13.8
IV	—	—	75.6	23.3	18.4	6.1
V	98.8	91.2	37.5	18.7	2.5	6.3
VI	72.5	53.7	6.3	0.4	—	0.8

an English university also), Groups V and VI consisted of girls in a secondary school, who had been doing English for about 5 years.

The test results show clearly the close correlation between the relative frequency of a word and its likelihood of being known by the students. Diagram 3, which is based on Table 7, exhibits the resulting vocabulary

DIAGRAM 3. VOCABULARY PROFILES OF SWEDISH STUDENTS OF ENGLISH, SIX DIFFERENT PROFICIENCY CLASSES (BASED ON TABLE 7). SEE TEXT, SECTION 8.



profiles of each proficiency group. The curves are extended leftwards and rightwards, where broken lines indicate their hypothetical continuations.

The next step is to calculate the aggregate vocabularies of each proficiency group by combining the data of Table 7 with those of Table 4.

TABLE 8. VOCABULARIES OF SEMANTIC UNITS IN SIX GROUPS OF STUDENTS WITH DIFFERENT STANDARDS OF PROFICIENCY.

p range down to	V in range	I		II		III	
		%	V	%	V	%	V
2×10^{-4}	640	100	640	100	640	100	640
8.2×10^{-5}	810	100	810	100	810	100	810
3.3×10^{-5}	1,050	100	1,050	100	1,050	100	1,050
1.43×10^{-5}	1,500	98	1,470	93	1,400	85	1,280
2×10^{-6}	5,200	90	4,680	65	3,380	40	2,080
6.7×10^{-7}	4,000	80	3,200	40	1,600	20	800
2×10^{-7}	4,500	60	2,700	25	1,120	15	680
5×10^{-8}	5,800	35	2,020	15	870	5	290
1×10^{-8}	6,500	20	1,300	10	650	1	70
Total	30,000		17,870		11,520		7,700
Corresp. theor. W	40×10^6		3×10^6		550,000		180,000
Corresp. average W (homogeneous)	—		—		—		1,800,000

IV		V		VI	
%	V	%	V	%	V
100	640	100	640	100	640
100	810	100	810	75	610
95	1,000	90	950	50	530
75	1,120	40	600	10	150
25	1,300	20	1,040	2	100
15	600	5	200	1	40
10	450	2	90	0	0
2	120	1	60	0	0
1	70	0	0	0	0
	6,110		4,390		2,070
	80,000		35,000		8,500
	650,000		220,000		30,000

The result of these calculations appears from Table 8. The percentages used in Table 8 are rounded off and somewhat adjusted to yield smooth curves for the (partly hypothetical) vocabulary profiles.

The values in the last two rows have been read from Diagram 2. They show, for instance, that the vocabulary possessed by Group III, about 7,700 semantic elements, could have been obtained from reading a text mass of 180,000 words (about 600 ordinary book pages) selected from widely different kinds of literature. The values in this row entered for the lower proficiency groups do not mean much, since the small text masses involved cannot possibly be sufficiently mixed. It is more realistic to consider the values entered in the last row, which shows the size of an average consecutive text likely to contain the vocabulary in question. No entries are made for the larger vocabularies, since the great text masses involved are necessarily diverse. The reasoning underlying this argument was set out above, section 4.

It must be stressed that the vocabularies entered in the "Total" row in Table 7 are not the total vocabularies of the groups, since they do not include the p ranges below 10^{-8}. On the other hand, if the vocabulary profiles we have drawn in Diagram 3 are approximately correct, the percentage of words known in the lower frequency ranges is so small that the additional vocabularies are not likely to be considerable, except in the very highest proficiency groups.[30]

The vocabulary profiles described above have been established for students learning English as a foreign language. It would be interesting to ascertain whether natives go through similar stages when acquiring their mother tongue, and whether the vocabularies of adult native speakers can be fruitfully described in similar terms. There is hardly any reason to suppose the contrary. The test tool, however, needs sharpening. What is required above all is more extensive, and more carefully constructed, frequency dictionaries.

Department of English Language and Literature
University of Gothenburg

[30] In strictness, of course, the W values in the last two rows should be those corresponding to the total vocabularies, not, as in the table, those corresponding to the vocabularies as curtailed above the frequency $p = 10^{-8}$. But the correction is of little practical importance.

Obituary

HENRI FRANÇOIS MULLER

PAULINE TAYLOR

Henri François Muller was born at Issy-les-Moulineaux, near Paris, on January 27, 1879. He received his early education in France, and in 1897 obtained his Bachelier-ès-Lettres degree. Coming to the United States shortly before the turn of the century, he began his university teaching career as a tutor in Romance Languages at Columbia, teaching partly in Columbia College and partly in Barnard,[1] while at the same time working for his doctorate, which he obtained in 1912. In 1909 he was made instructor, and in 1914 was promoted to assistant professor. When the First World War broke out, he obtained a leave of absence from the University in order to serve in the French army, and during the latter part of the war acted as liaison officer with American General Headquarters. Returning to his teaching post at Columbia after the close of hostilities, in 1924 he was promoted to associate professor of Romance Languages, and in 1927 he was made Professor of Romance Philology, a post he held until his retirement in 1944. From 1929 to 1936 he served as Executive Officer of the Department of French and Romance Philology. In 1929 the French government honored him for his services to French culture by bestowing on him the Croix de Chevalier de la Légion d'Honneur.

He was a member of numerous scholarly societies, among them the Modern Language Association of America, the American Association of University Professors, the Société des Professeurs Français en Amérique, the Société des Etudes Latines, the Société de Linguistique de Paris, Vox Romanica, and the Zeitschrift für Romanische Philologie. He was one of those who, in 1943, founded the Linguistic Circle of New York, and he served as president of the Circle from its beginning until 1956. For

[1] The author of this article, (Professor of French at New York University), studied first as an undergraduate in Barnard College with Professor Muller. Later, as a graduate student, she became his first candidate for M.A. and Ph.D. degrees in Columbia University. In 1958, while doing research in Europe under a second Fulbright grant, she was invited by the University of Louvain to lecture on Muller's theories concerning Vulgar Latin.

twenty-two years he was a member of the editorial board of the *Romanic Review*, and for many years he served in an editorial capacity on *Word*, *Vigiliae Christianae*, and *Le Français Moderne*. He was also on the editorial board of the *Todd Memorial Volumes*. After his retirement from Columbia, he lectured for some years at the Ecole Libre des Hautes Etudes in New York City and at Laval University in Quebec.

His courses at Columbia included Vulgar Latin, Romance Philology, Old French, the history of the theatre in France, and the Renaissance in France. His large classes were an ample testimonial to a brilliant and scholarly personality which gave life to words and happenings of a remote past, showed the inter-relations of language, history, literature, and human beings, evoked far-away centuries, and animated everything he discussed. He was always accessible to his students, spent unending hours working on their problems, was kind, sympathetic, encouraging, and inspiring. At various social functions at Columbia his songs entranced his audiences, his skill as an actor was acclaimed, and his talks were always popular. The rare qualities of humanity, sympathy, and charm combined with profound erudition endeared him to all who studied and worked with him.

During the last five years of his life, Professor Muller retired altogether from scholarly and social activities. His death, which occurred on September 10, 1959, leaves his former colleagues and students, as well as all others who knew him, filled with a sense of profound bereavement.

Starting in 1912 with his doctoral dissertation (*Origine et Histoire de la préposition à*) in which he first stated his belief in the intimate relation between the written and spoken language in Gaul up to the end of the eighth century, continuing this hypothesis in 1921 (*When Did Latin Cease to be a Spoken Language in France*), reaffirming his conviction in 1923 (*On the Use of the Lingua Romana from the First to the Ninth Century*), in 1924 (*The Passive Voice in Vulgar Latin*), and finally, presenting his views as a whole in 1929 with the *Chronology of Vulgar Latin* which gave his definitive statement, Mr. Muller constructed his *monumentum aere perennius*. In 1932 the *Chrestomathy of Vulgar Latin*, produced in collaboration with Pauline Taylor, permitted the practical application of this new interpretation of Vulgar Latin and his cultural interpretation of the Merovingian period. The extraordinary intuition, the penetrating insight brought to bear in his examination of hundreds of documents and studies in the so-called Dark Ages have illumined this nebulous, misunderstood, neglected, and all-important period of pre-Romance philology and established a knowledge of the period as an indispensable requisite for the study of Romance philology.

The traditional interpretation of the linguistic aspects of the sixth,

seventh and eighth centuries can be briefly summed up (Mohl, Grandgent, Meyer-Lübke, Brunot, Bourciez, Nyrop et al.): The documents of these centuries are full of gross errors resulting from the ignorance of Classical Latin on the part of the writers of these periods who were trying to write correctly, but were influenced by the vulgar or Romance tongue which they spoke and which was spoken around them. The texts, filled with illogical errors, do not represent the spoken language which was Romance, already formed. Regional features, Romance dialects were already in existence.

Mr. Muller, examining innumerable documents in these centuries, proved by the facts found therein that these so-called gross and illogical errors, tending in regular, definite directions, increasing in frequency up to the end of the eighth century, were, in fact, the manifestations of the new Romance tongue which was to appear in the ninth century; they were pre-Romance phenomena to be interpreted accordingly, not only in the light of past Classical Latin, the mother tongue, but rather in the light of the new language of the future to appear. Mr. Muller demonstrated that the scribes wrote what they spoke and heard, Latin (Vulgar Latin, Folk Latin, bas-latin, latin rustique, Merovingian Latin, sermo vulgaris, lingua romana, lingua romana rustica, etc.), i.e. the language of their country. He found no significant differences in the Latin throughout Romania up to the ninth century and therefore no evidence of dialectalization or the formation of Romance languages before the ninth century.

The wealth of new pre-Romance linguistic material in these centuries recognized as such by Mr. Muller and his students[2] can be mentioned only in part:[3] the transformation and reduction of the Latin word by the dropping of the unaccented medial and final syllables (*colocare* > *culcare*, *locus qui dicitur Camp* < *Campus*), the production of a two-case declension tending towards a one-case declension (*murus, i, o, um, o* > *murus, o* > *muro*), the creation of the article (*et illa alia medietas cleri illas veteres cappas quas illi seniores*, etc.), the creation of the future (*Diabolus habet*

[2] Taylor, Pauline. *The Latinity of the 'Liber Historiae Francorum'*, New York, 1924; Pei, Mario A. *The Language of the Eighth-Century Texts in Northern France*, New York, 1932; Trager, George L. *The Use of the Latin Demonstratives, especially 'ille' and 'ipse'*, New York, 1932; MacGonagle, Sara H. *The Poor in Gregory of Tours*, New York, 1936; Sas, Louis F. *The Noun Declension System in Merovingian Latin*, Paris, 1937; Scott, Ralph W. *A Study of the Uses of the Prepositions 'ab', 'apud' and 'cum'*, Tokyo, Japan, Maruzen, 1937; Jennings, Augustus C. *A Linguistic Study of the 'Cartulario de San Vicente Oviedo'*, New York, 1940.

[3] Cf. *L'Epoque mérovingienne*, chapters IX–XVI.

quaerere causas), of the factitive locution (*Ad bonis hominebus atfirmare rogavimus*), the transformation of the passive, the beginning of the use of *homo > on* for the Latin passive, of compound tenses (*audivimus > auditum habemus*), the increase in the use of prepositions, the formation of prepositional and adverbial phrases (*in rotundo > en rond*), the increase in vocabulary (*ficato > foie, granica > grange, carrada > charrée, intaliare > entailler, vindicare > venger, werra > guerre*), the beginning of the new Romance order of words in a sentence.

Mr. Muller pierced through the darkness of an apparently confused, chaotic, fluctuating and therefore incomprehensible mass of words teeming with orthographical, morphological, and syntactical heresies[4] committed against Classical Latin, and discerned therein a groping in the early centuries in certain directions, a more definite movement as time progressed and finally at the end of the eighth century a new grammar, a new language, we might say a new order. Often the new pre-Romance construction would be found side by side with the more conservative Classical, often the changed and reduced morphological endings of the new case system would be found with the former Classical forms. But the new phenomena multiplying in the documents indicated slowly but surely the formation of the pre-Romance tongue which, with the epoch-making Edict of Charlemagne in 813 demonstrated the separation of the daughter, Romance, from her mother, Latin. And so was born the new language of which we have the first signs in 842 with the Oath of Strassburg. And we might add that, in general, Mediaeval Latin, a pseudo-Classical Latin, can be assumed to function (at least in Gaul) after the reform of Charlemagne and was used for learned literature and by the ecclesiastical world.

A delicate question, and one on which Mr. Muller has shed considerable light in the *Chronology of Vulgar Latin*[5] is the dearth of Vulgar Latin documents in Italy and Spain. It was France alone in which we find this extraordinary abundance of material. Italy possessed a highly cultivated aristocracy in which the adherence to the Classical Latin functioned to maintain the purity of the language and to discourage writing in the Vulgar tongue. This resulted in the first mention in Italy of the Vulgar language a century later than in France (Monte Cassino, 960). In fact, throughout the Middle Ages the Vulgar (as distinguished from serious writings in Latin) literature of France is an ever-fresh fount of literary inspiration to the

[4] The writer of this article, on showing Merovingian documents to Classical scholars, has been struck by the unconcealed disgust (and lack of comprehension) at this offensive mess: *extiblatione* for *stipulatione, puella qui* for *puella quae*! Indeed!! They refused to look further.

[5] Chapters XV–XVII.

Occident. In Spain similar aristocratic conditions did not permit signs of the vulgar tongue until the tenth century.

In his *Epoque mérovingienne*, Mr. Muller has, indeed, concentrated practically exclusively on Gaul.[6] He lifted the curtain from the so called Dark Ages there and proved this Merovingian period to have had a remarkably vital, creative activity. He showed how, for the first time in Romania and through the psychological influence of Christianity, the dignity of each man, possessed of an immortal soul, produced the triumph of the masses away from the impersonal Classical Latin to the personal Vulgar Latin read constantly to all in homilies in the language they knew and spoke. The accent of intensity, syncope, labialization, palatalization, nasalization, the use of the article, the development of the pronoun, the reduction in declensions, the changes in compound tenses, the new passive voice, a new vocabulary were the outward manifestations of a new personal, Christian language. He showed how, particularly in Gaul, the increasingly important bishops were the protectors of the poor, how they created the conscience of the people, how the monks organized agriculture, industry, commerce, and art, and built churches and monasteries which became the center of the spiritual and material growth of the people.

The overwhelming mass of facts and statistics produced by Mr. Muller may be said to have forever dispelled the old belief in a static, moribund society of the sixth, seventh, and eighth centuries. His *Epoque mérovingienne* brings to mind *Le Génie du christianisme* of Chateaubriand or the extraordinary evocations of *La Légende des siècles* of Hugo. Since, in the Merovingian period, is to be found, by a study of data, the beginning of the formation of the Romance tongues, of the popularization of the arts, sciences, and jurisprudence, of the ecclesiastical calendar, of socialized Christianity, the study of this period must form a vital part of the study of Romance languages, history, the church, the arts, in fact, all studies based on or including the history of Western Civilization.

Mr. Muller's new interpretation of Vulgar Latin phenomena has met with many objections on the part of those scholars who hold to the substratum theory, who believe in early dialectalization, who ascribe less linguistic importance to the Christianization of Romania, who believe in the Age and Area hypothesis, in the Germanic influence on pre-Romance in Gaul. Where Mr. Muller, however, has differed from his predecessors and contemporaries in his understanding of Romance Philology, is in his having focussed their attention on the only known facts concerning the

[6] Chapters V–VII.

Merovingian period, those found in the documents, his having established the development of many Romance phenomena through the study of these documents, his having proved the existence of the wealth of information to be found for the linguist and historian in the once-called Dark Ages.[7]

New York University

[7] The following quotations indicate the general change in attitude of historians and philologists towards this question: A. Burger in *Mémorial des Etudes Latines*, Paris, 1943, page 163, "C'est la langue parlée sans doute, mais parlée par tout le monde, c'est le *sermo vulgaris* de l'époque mérovingienne;" page 165, "Le roman commun s'étend, en gros, du IIIe au VIIIe siècle; il s'élabore à la suite de l'expansion du latin dans l'Empire et du fléchissement de la culture antique; l'évolution indépendante des diverses langues romanes résulte de l'invasion arabe et de la ruine du commerce méditerranéen comme l'a bien vu M. H. F. Muller." Henri Pirenne in *Mahomet et Charlemagne*, Paris, 1937, page 122, "C'est ce que les lettrés appellent le latin rustique. Mais ils s'y prêtent et l'emploient surtout en Gaule, parce qu'il est la langue populaire, celle de tous;" pages 251–252, "Ceux qui le (le latin mérovingien) parlent, peuvent comprendre et se faire comprendre partout dans la Romania . . . . L'église n'hésite pas plus à s'en servir pour ses besoins de propagande que l'administration et la justice . . . le fait certain est que le latin comme tel a cessé d'être entendu vers 800 sauf par le clergé." Ferdinant Lot in *Histoire du moyen-âge*, vol. I, Paris, 1928, page 612, "Le latin parlé avait prodigieusement évolué depuis la fin de l'Empire romain. Or les hommes instruits, même dans le clergé le prononçaient comme la masse de la population et ils l'écrivaient à peu près comme ils le prononçaient". (On this change of opinion in Lot and the date on which this statement was written cf. Augustus C. Jennings, *A Linguistic Study of the 'Cartulario de San Vicente de Oviedo'* (New York, 1940), page 312.)

Review Article

LINGUISTIC STRUCTURES

W. Haas

After years of vigorous development in linguistic studies, there seems to be a feeling abroad amongst American linguists that the time has come for taking stock, and even, once more, for the writing of comprehensive textbooks. We have had three such books in quick succession: *An Introduction to Descriptive Linguistics*, by H. A. Gleason in 1956, *A Course in Modern Linguistics*, by C. F. Hockett in 1958, and, finally, the present *Introduction to Linguistic Structures* (*From Sound to Sentence in English*), by A. A. Hill.[1] These three studies have had their scope defined in different ways; but they naturally overlap over wide and central regions. Hockett's study, as is indicated by its title, is the most comprehensive; and the two works which have limited themselves to descriptive linguistics differ in intention.

Hill's book, as the *Foreword* informs us, is intended to serve two purposes at the same time—on the one hand, as a text for a course in Descriptive Linguistics (which is also what Gleason aimed at), on the other hand, as a text for a course in the English language. This combination of two tasks may, at times, try the patience of either of the two types of reader envisaged by the author: the student of Linguistics may feel that he is given more about English than he requires for his purpose, and the student of English that he is asked to give too much attention to a "step-by-step exposition of analytical procedure" (VI). But if patience and perseverance are needed from both in taking in more than is strictly required, neither can complain that he is given more than is good for him. The English-speaking student of *Linguistics* can hardly do better than apply the general techniques of his discipline to a detailed study of his own tongue. The student of *English* should welcome the opportunity of following the actual course of analysis, which yields the description of the language he is interested in; he might well agree with the author that "the results of linguistic investigation since 1933 have been so considerable" that the general task of "redrawing the broad outlines of language study on newer lines" (VI) had to come first even for the purpose of a description of

[1] Archibald A. Hill, *Introduction to Linguistic Structures: From Sound to Sentence in English*, 496 pp., New York, 1958.

English. The question is really not, whether the author has given us too much but only whether he has given us enough to satisfy both types of interest.

Whatever his attitude to the general plan of the work, the student of English will find that its 500 pages of close print are devoted almost entirely to providing him with a full and meticulous analysis of the English language—and an analysis which is worth his attention, because it is, in many ways, new. Even the structural sketches of the two appendices—some twenty pages on Eskimo and forty on Latin—are intended to throw "a little further light on the structure of English" by contrast with these other languages. The two sketches are of course intended also further to illustrate "techniques and attitudes which should be useful for all languages, or for none" (418). However, what the student of General Linguistics may miss in the main body of this book is not so much an application of general techniques to non-English material as rather a clear and connected statement of these techniques. Methodological remarks are dispersed throughout the text. Many appear in the form of justificatory asides, and it is not always easy to see their systematic connection. The work as a whole will have to be regarded as primarily a massive contribution to the study of English. But Professor Hill does adhere to a clear-cut procedure of linguistic inquiry; and for a general survey of the work, it might perhaps be best to view it as an application of that general procedure. This would seem to meet the author's expectations.

I am in fundamental disagreement with Professor Hill. It is therefore right to say at once and emphatically, that even if all the objections I find I have to make, to Professor Hill's general procedure, were considered valid, they could not detract from the excellency and usefulness of many of the specific results which he has obtained by applying that procedure. This procedure may be found to break down at critical points, but it certainly does not reveal its weakness at every point.

Completeness, Consistency, Simplicity

The significance which attaches to any work of stock-taking is that it reveals gaps. Having undertaken to proceed step-by-step "through the hierarchy of English structure, from the smallest elements, sounds, up to the largest elements, sentences," Professor Hill has aimed at giving "a balanced amount of attention to each level, so that the result is unified" (p. V). In this, he may rightly claim the book "differs from previous attempts to apply newer techniques to English"—most notably from C. F. Fries' *The Structure of English* which deals with syntax only, and from the Trager-Smith *Outline of English Structure* which, though most congenial

to Professor Hill's approach, is primarily a treatment of sounds. His own main task, Professor Hill says, has been a work of "unification"—an attempt, then, to fill the gaps.

There are various methods of filling gaps. One is the *eclectic* method. It consists in freely multiplying and mixing principles of analysis. Whenever one approach is found wanting, another is adopted. An example is Mr. Gleason's procedure. It consists mainly in giving American linguistics an occasional sprinkling of Glossematics, and offering alternative definitions of fundamental terms (unless, indeed, they are left undefined). This is a way of securing empirical completeness at the expense of consistency. Such a book can be a lively help for teacher and student—and this is one of the great virtues of Gleason's book—but it will hardly advance linguistic thought. It will be apt to stimulate the student's interest, but at the risk of blunting the point of his theoretic energy.

Another way with gaps is to fill them in an *informal* manner, as did Bloomfield for example, when he confessed to having no precise technique for dealing with meaning, and yet proceeded to "presuppose a knowledge of meaning" (*Language*, pp. 137ff.). Hockett, in his comprehensive *Course*, proceeds in such fashion. It is a kind of approach which aims to be empirically adequate—keeping in touch with the rich concreteness of our experience of language—*without* offending our intellectual conscience. The *system* of description is explicitly declared not to be exhaustive; it relies on information drawn from outside. As far as it goes, it remains consistent, and free of theoretical confusion. But we are left with some loose ends. This will do no harm, so long as we preserve our theoretic ambition, i.e. continue to regard the incompleteness as a challenge, and every informal bit of outside information as a demand for formal inclusion.

No one could accuse Professor Hill's analysis of showing "loose ends." For him, *completeness* of systematic description is the most important of the three criteria by which he judges the "rightness" of an analysis (48, 53). Only in the extreme case will he permit himself to say either that reliable data are not available, or even let himself go so far as to say that he has no precise method of dealing with the facts. Generally in such a case, he prefers to leave the gap. He will conscientiously mark it, rather than offend against *consistency*—this being his second criterion of adequacy. Real gaps in his description are few. The question, however—as in the case both of the eclectic and the informal remedy—must again be: at what price? If his treatment is as complete as the eclectic and as consistent as the formally incomplete, will it then only be at the expense of *simplicity* that he allows himself to deal with the exigency of gaps—'simplicity' being his third criterion? This would, perhaps, be more than we can rightly expect.

We find, I think, that a higher price has to be paid. As Professor Hill aims at making the framework of his system inclusive, he seems to be forced, at times, to put up with worse than a complicated description: he has to tolerate, and he must ask us to tolerate, a certain amount of distortion as well. What is complicated, elusive, and insignificant will at times be substituted for what is in fact simple, obvious, and important. As the chosen principles are kept pure, and their exhaustive adequacy is insisted on, some 'hocus-pocus' appears to be unavoidable. When the interpretative principles are not quite adequate, and are yet not allowed to yield, the facts have to. But this, precisely—irritating though it must be at times—is the reason why this is a book of great theoretical interest. Nothing is more stimulating for the advancement of linguistic thought than this constant invitation of clash between principle and facts. There can be no better test of principles of analysis than an application of them so extensive and rigorously consistent as not even to flinch from absurd consequences. As the author says: "In a structural description, rigorous analysis occupies much the place of experimentation in laboratory sciences" (p. vi). The experiment will indeed be rather hard going for the student who is supposed only to be *introduced* to linguistic structures; and for the student supposed to be introduced to the structure of English, even harder. But anyone who approaches this book with a certain fund of experience in the two fields of study, will recognise the theoretic passion behind its rigour, and appreciate the intellectual energy behind its meticulousness and self-imposed restrictions.

A Minimum of Meaning

What then are the principal theoretical restrictions which the author has chosen to impose upon himself? As one may have expected after reading his acknowledgements in the *Foreword*, they concern the problem of meaning. It all looks simple enough, at first. "The description and analysis of language must begin with description of the sounds and their patterning." If we wish to proceed "from the more knowable to the less knowable," then "description of meaning must be put off" (p. 3).

These remarks might suggest that Professor Hill is proposing to carry out the whole task of linguistic analysis without referring to meanings at all. The very first pages of the following phonological analysis (pp. 15ff.) raise doubts of this. They are crowded with references to semantic contrasts. Are these contrasts, then, no part of a "description of *meaning*?" Or are they, contrary to appearance, not really part of the *linguistic* description? The puzzled student must wait for an answer. On p. 90, he will read that

"analysts . . . make use of meaning in spite of the danger." What they are warned against is only "reliance on meaning as a primary tool of analysis." In dealing with the elements of language, "identification and classification must rest firmly on formal and distributional characteristics rather than on meanings." What then *is* the role of meaning within linguistic analysis? Some kind of an answer seems to be given further on: ". . . even if an analyst has used meaning as an *initial guide*, he should *back up* his decisions by distributional statements. Meaning should be no more than *a hint of what to look for*" (p. 94). No part then, of the final linguistic description? No more than a preliminary heuristic help? This seems clearly suggested. But it is contradicted throughout the book. The distributional statements which are intended to "back up" the linguist's guiding intuitions of meaning, contain themselves constant references to meaning: to "sameness" or "difference" of meaning, to what is a "possible" or "impossible", a "normal" or "unnatural" utterance—this, as an integral part of the whole "backing-up" operation. *Some* references to meaning seem to be centrally involved in linguistic description.

To cite a few examples: A central notion in the phonemic classification of sounds is 'contrastive distribution'—a notion which implies "*difference in meaning*" (pp. 50ff.). Again, morphs—like 'yes' and 'no' for instance—are said to be members of different morphemes, if they contrast in the same environments (p. 96). Similarly, "*sameness of meaning*" is accepted as of central importance; for establishing 'free variants' in phonology (p. 50), or morphology (pp. 92, 97, 117); or again, for recognising "possible" junctures in fixing the boundaries of 'phrases' (pp. 124, 185), 'sentence-elements' (pp. 201, 259), sentences (p. 363); or even for recognising a 'part of speech' ('sentence-element') in particular forms (e.g. by free variation of verb forms, p. 274, or of order, p. 333).

And there is one further type of semantic statement, also appearing as an integral part of a structural description. In order to see that it is a further type, we need only realise that to say two utterances do *not differ* in meaning does *not* amount to saying that they have the *same* meaning. Both utterances, or one of them, may have no meaning at all. The third type of semantic statement we require in linguistic analysis is implied by the two others, but says less; it is to the effect that an utterance 'has meaning' (no matter what meaning). In the many cases, in which we are told, for purposes of structural definition, that a certain element *can* be replaced or omitted, or that an utterance *can* be expanded or otherwise transformed, we are obviously supposed to recognise that these operations result in meaningful and correct utterances; whereas in other cases (in those expressions which are marked with an asterisk) they don't. Examples

are the definitions of noun-phrase (p. 175), verb-phrase (p. 218), determinative phrase (p. 187) adjective (pp. 179, 186), or again, decisions as to an element's class-membership: noun *versus* adjective (p. 176), adjective *versus* participle (p. 203), conjunction *versus* adverb (p. 400), and many more.

It is strange that these statements about the 'possibility' or 'impossibility' of utterances are not explicitly recognised as implying semantic considerations; even more strange that they are sometimes explicitly declared not to imply such. (For example, having made use himself of a "test of expandability" which requires him to distinguish possible from impossible expansions, Professor Hill tells us that "throughout" there has been "no attention to meaning or function" (pp. 123, 203, 205). Such a claim might conceivably be justified, if we confined our description to a finite text, and if instead of asking whether an utterance was "possible" or "impossible," we confined ourselves to stating that certain utterances did in fact occur in our text, while others (though, perhaps, possible) did not. Description of a 'dead' language might conceivably work under such restrictions. I am not acquainted with any example. Even here, it is surely general practice to go beyond statements of actual occurrence, and to 'predict' (as it is the fashion to say) what was possible. To treat a text as if it were incapable of being extended would be to ignore the essential 'openness' of language, and therefore to give an incomplete and distorted account of it. Being concerned to state what restrictions are imposed on any speaker's freedom of utterance, Linguistic Analysis must be concerned with 'possibilities'—i.e. possibilities of meaningful utterance.

How does Professor Hill succeed in hiding from himself this minimal implication of meaning? Could it be by the mere avoidance of the word? He tells us: "Statements that so-and-so does not or cannot occur mean that the informant rejects them as foreign to his habits." And also: "I have had to use myself as the informant" (p. 13). This is, of course, quite unobjectionable. What the linguist can tell us about his own habits is no worse than what other informants might have told him about theirs. But, those habits? We ought to be more explicit about them. *What* habits? Obviously concerning the use of linguistic material—a habit, mainly, of accepting uses which "make sense" and rejecting others. If 'correctness' were here appealed to, rather than 'meaning', this would ask for more rather than less. For any 'correct' utterance, even the most abstruse—such as Mr. Chomsky's "Colourless green ideas sleep furiously" or Bertrand Russell's "Quadruplicity drinks procrastination"—is meaningful at least potentially. It carries a presumption of meaning, which can always be realised in a suitable context.

The first fleeting indication of the author's real position, with regard to the cardinal problem of meaning, may be found on p. 95, when he speaks of "*minimising* the reliance on meaning." For a fuller statement we must turn to the very end of the book. There, we learn that, within Linguistic Analysis, "only what has been called '*differential meaning*' is relevant" (p. 409). This minimum of 'microlinguistic meaning' (as G. L. Trager called it) seems to be incapable of being exorcised. The whole book provides illustrations of this fact; but it is only on the last pages that Professor Hill comes near to explaining, how exactly linguistic analysis takes account of meaning. The crucial passage, as simple as it is fundamental, runs as follows:

> "The identity-difference which constitutes microlinguistic meaning can be studied by the process of observing whether substitution of one item for the other produces *identity or difference in a larger structure*. That is to say, in studying phonemics, we substitute one sound for the other and observe whether the substitution produces identity or difference in morphemes. In studying morphemics, we substitute entities which produce identity or difference in words and fixed phrases, and so on through the hierarchy" (p. 410).

Clearly, the 'identity or difference', which may be produced in the 'larger structure' by replacing a part of it, must be identity or difference in *meaning*. As to its *form*, the whole is bound to be different if we replace a part; and nothing could be studied by 'observing' a foregone conclusion. It follows that what is meant, here, by 'a larger structure' is always a *meaningful* structure, and that criteria or plain intuitions of "having meaning" are presupposed by the 'substitution-test'. Could we assume that Professor Hill has omitted to mention these implications, because regarding them as too obvious? There are doubts about that. Here is the place in linguistic theory which requires to be filled by a minimum of semantic intuition. Bloomfield expressly acknowledged this 'gap' in the theory. Professor Hill seems to be strongly tempted to plaster it over. "Differential meaning," he says, means "no more than identity and difference." In linguistic analysis ('microlinguistically') we are supposed to be able to say simply "that two items are 'sames' or 'differents' and nothing else" (pp. 409ff.). This, surely, wont do: some distinction between meaning and form, and between 'differential meaning' and 'differential form', needs to be made. Throughout the hierarchy of language, what counts in substitution-procedures of analysis is 'commutation' (to use Hjelmslev's term), i.e. mainly, a correlation of *form-differences* (between exchanged constituents) with *meaning-differences* (between the larger structures).

To obscure the way in which, here, considerations of meaning enter linguistic analysis, is detrimental to the analysis. For the study of meaning it is fatal. This is how semantic theory is being deprived of what must be its most promising growing-points. Those who seek to make semantic studies lose their foothold within the field of linguistics, do in fact promote the very kind of uncontrolled and obscure semantics which they are trying to avoid. It is unfortunate that, in recent American linguistics, the term 'semantics' has come to be used exclusively for non-linguistic treatments of meaning, most of them vague and anecdotal, while all sorts of substitutes—native response, habit, hearer-speaker-behavior, etc.—are used to hide the implication of meaning in linguistics. Professor Hill describes as 'semantic' only such of his observations as constitute optional additions to his analysis. Moreover, though helpful on the whole, they are supposed to be about a dubious kind of entity, 'semantic components' (pp. 205ff., 219–220); while the semantic characteristics of plain *linguistic* components—the distinctive values they have among others, the roles they play within larger structures—are not recognised for what they are.

The 'meaning' of a linguistic expression is still supposed to be dealt with primarily by definition, conceptual or referential—that is, either of an idea in the mind or of a range of external things (pp. 90, 152, 410). We may well agree with Professor Hill that this way of treating meaning is only of marginal interest for linguistic analysis. We may think the same of attempts to define 'semantic components'. The semantic value of an expression, on the other hand, viewed as its *uses*, its 'privileges of occurrence', *is* the linguist's concern, and is of central importance for him.

Only the very last section of the book (not more than four pages) deals explicitly with linguistic meaning. We are given some indication, here, of how it can be 'backed up' and controlled. "Meaning," we learn, "can be defined as partial predictability" (p. 413). The link with 'commutation' is obvious; though, curiously, it is not mentioned. For the point is that a contrastive element which occurs within the frame of some larger structure, and is *itself not* determined by the frame (i.e. which is to that extent unpredictable), may be looked at as chosen from a certain *range* of elements which *is* so determined. As long as the frame is not larger than a sentence, we are said to be dealing with meaning on the 'linguistic' level; if we expand the frame to a larger piece of text, or even to the context of a situation, we would be said to be moving on to 'stylistic' and 'extralinguistic' considerations of meaning (p. 414). Professor Hill does not say it, but these are obviously the facts which are investigated by the procedure of contrastive substitution; and this is how linguists who apply the procedure "make use of meaning" as "a hint of what to look for." What can be regarded as

chosen from determined ranges of linguistic items is linguistically significant ('non-redundant') and—as we should add, though it is a point often ignored—what *regularly determines* such specific ranges of choice is also important (even if it has little contrastive value or none at all).

Ever since Saussure, linguists have been clearly aware of the central importance of these techniques. The question arises why, more recently, some American schools of linguistics have been feeling uneasy about them; why Professor Hill, for instance, though using them sporadically throughout his analysis, and otherwise intent on giving a "step-by-step exposition" of his procedures, should have kept this one obscure to the last. The answer seems to be that a timely and full appreciation of it would have threatened his whole programme of analysis.

From Sound to Sentence in English

—this is the programme, as announced by the sub-title of the book. It is declared clearly to be a programme of *analysis*, not merely an 'order of presentation' (pp. 13–17). The techniques by which "microlinguistic meaning can be studied," and needs to be studied, are diametrically opposed to it. At every level, as we try to make the proposed ascent through the 'hierarchy of language', contrastive substitution is one step ahead; it implies that we are already acquainted with units of the *higher* levels. Substitutions are made within the frames of 'larger structures'—in 'functions' (to use the terminology of logic). Sounds are found to contrast in morphological functions, morphs in lexical or phrasal functions, and so on. The same requirement holds for the investigation of syntagmatic dependencies; they can only be stated *within* determinate structures. The use of these analytic techniques would accord with a *reductive* order of inquiry, from sentence to sound; but they contravene Professor Hill's *expansive* order at every step.[2]

It is not necessary, of course, and does not seem to be advisable, that a reductive order of analysis should be adhered to throughout. Necessary, is only that *some* larger structure should be available at every successive level. This is assured, if we *begin* with sentences. The decision that, for the

[2] I am using the terms 'expansive' and 'reductive' for the two kinds of procedure—in preference to the distinction between 'working up' and 'working down' through 'the hierarchy'. The latter terminology would suggest that the opposed procedures share one and the same notion of linguistic hierarchy—which, as we shall see, is patently false. The so-called hierarchy of language is not just 'given'. It is the product of linguistic analysis; and each of the two procedures produces another. Either may be said to 'work up' to *its* top-level. The decisive question is what to take for the ground-level of the hierarchy—sentences or sounds.

purpose of a strictly linguistic analysis, we do not go *beyond* the sentence, but *begin* with it, immediately commits us to a minimum of primitive semantic intuitions—namely, *that* sentences have meaning (no matter what) and that they *differ* in meaning. With this, we can move freely from level to level within the sentence, and substitution-techniques can provide us with a test of relevance throughout.

On the other hand, if we tried to keep strictly to Professor Hill's programme—ascending from sound to sentence, and never looking at any higher level before we have reached it—then, we might travel but surely could never arrive. Reductive analysis has arrived from the start; it looks back at familiar country—total meaningful utterances—and reduces it to a linguistic map. By the expansive procedure on the other hand, we are supposed to reach lands unknown—and this without a map, even without being allowed to look ahead. It would only be by keeping our eyes turned backward that we could claim to be advancing to higher levels of analysis. This is the paradox of trying to analyse language while building it up.

Professor Hill, we have seen, does not keep strictly to a backward-looking procedure. In order to climb his hierarchy, he is prepared at times to look ahead for the right direction. "The method of work," he says at the very beginning, "has been to try to group my utterances into those which are the same and those which are different," and "attention has been focused on the differences—the contrasts" (p. 14). We have here, it is true, the fatal ambiguity of 'same' and 'different', and also of 'utterance'. Nothing yet of *meanings* being same or different, and nothing of *larger structures* of utterance. But we have seen that, in fact, neither of these two is "put off." In fact, they are made use of at every stage. How is this to be reconciled with Professor Hill's programme of analysis? How does he justify a 'method of work' which is always at least one step ahead of schedule? He does not tell us, and we can only guess. But his remark on using meaning as an informal "hint of what to look for" suggests what his justification might be. In some of its applications, at any rate, Professor Hill would seem to regard his 'method of work' as something like an informal trick of the trade; a device for obtaining the occasional glance forward which is needed for making progress in the formal backward-looking way. That is to say, reference to his 'higher' levels would be confined to playing an informal 'catalytic' role; would serve to *assist* the process of description, to start it or to speed it up; but it would never enter its end-product. The *description* or *definition* of a linguistic unit would then be given in terms of its *constituents*, not in terms of its *functions*.

Can even this lesser restriction really be upheld? Can functional state-

ments, if not excluded, at least be kept to a purely catalytic role? We have seen that a similar restriction cannot be imposed on references to meaning. It is not surprising then to find that it cannot be imposed on references to 'larger structures' either. For it is precisely by referring to an element's occurrence in 'larger structures' (to its 'functions') that we refer to its semantic value.

The catalytic view of the role of meaning and function, in linguistic description, may have a certain plausibility, so long as we are in the early stages of the expansive, combinatorial ascent 'from sound to sentence'. For example, at the phonemic level, we might claim to be using contrastive substitution only for selecting the relevant phonetic units, while their formal description was given in terms of their constituent articulatory features. Or again, at the beginning of morphological inquiries, the contrasts and syntagmatic relations of whole sequences of phonemes may be said to be used merely for the purpose of picking out the morphologically relevant amongst them, the morphs; while the formal description of the selected morphs is given in terms of their constituent phonemes. Such descriptions would be precarious, and not very informative. But even if we found them wholly acceptable, we should have to admit that we can go no further in this way. In order to be able to say, for instance, that the *-s* of *cats* is 'the same' as the *-es* of *horses*, or that the *yes* of *yesterday* is not 'the same' as that of *Yes, Mr. Jones*, the description itself of what is or is not the same must take account of the similarity or dissimilarity of the 'larger environments' and 'constructions, which contain the morphs in question' (pp. 96–97). Similarly in all the cases already mentioned where the various operations of substitution, expansion, omission are performed on *functions* of linguistic units, i.e. on structures which contain them; and where the units are not merely selected by these operations, but are *characterised* by the corresponding distributional relations. They are formally described as occurring in the type of function which admits, or does not admit, of those operations. Here belong also those 'order classes' which are defined with reference to a 'model-phrase' (pp. 175ff., 230ff.); or the extension of formally established 'core-groups' (of 'adjectives' or 'adverbs', pp. 167ff.). Professor Hill speaks of units "defined by their syntactic characteristics" (e.g. pp. 180, 218, 240) or by their "distributional characteristics" (e.g. pp. 168, 173).

Students of General Linguistics may well be puzzled about the status of these 'characteristics' in Professor Hill's general plan. For to introduce them is clearly contrary to it, and there is nothing one can do about it. To define a linguistic unit by its functional relations (by distributional relations in larger structures) would remain contrary to the expansive

(combinatorial) procedure of analysis, even if one chose to introduce such definition by way of subsequent revision—the familiar 're-phonemicising', 're-morphemicising', etc. This would assure that the larger structure is introduced according to schedule; but the revision itself, in defining a unit by appeal to units of higher level, would still run counter to the general direction of the analysis. Professor Hill occasionally has recourse to similar revisions and calls some of them, strangely, "removal of ambiguities" (pp. 182, 203). Ambiguity of what?—of the text, or of a previous analysis of it? But generally, in such cases, he does not worry much about the schedule. That is, if there is no alternative, he boldly and immediately refers to relations in 'larger structures', even if the latter are strictly "out of order." The need arises often enough. Deviation from his general programme of analysis is almost a regular feature in the later stages.

What, then, is the practical effect of the programme? Is it effective at all —since it is *not* to exclude techniques which work in the opposite direction, not even always to relegate such techniques to a purely catalytic role? We find that it is. Indeed, our main criticism of Professor Hill's description of English will be that those semantic and functional criteria which we have seen he cannot entirely avoid are yet not nearly given their proper weight. The decision to proceed 'from sound to sentence' is against it. For the decision does provide a powerful directive for the whole course of the analysis. Its effect has been, on every level, to reduce appreciation of functions to a minimum. Only as a last resort, a unit may be defined or 'characterised' by its relations 'in larger structures'. What for European structural linguistics—from Saussure to Hjelmslev—has been the very hinge of linguistic structure: namely, the 'syntagmatic' and 'paradigmatic' relations of items in their functions—is here reduced to a subordinate role. We are supposed to avoid using relations for defining their terms. In describing or defining any linguistic unit, we are asked to refer primarily to its *parts* and *their* relations, not *its* relations, *as a whole*, to other units. We are asked to examine primarily the parts it contains, not the "parts it plays." If at a later stage we happen to come upon its relations, the unit itself is supposed to have been described or defined already and without reference to them. What is laid down by the programme of expansive analysis is a strict *priority* of description in terms of constituents, over description in terms of functions—a priority, as one might say, of *componential* description over *functional*.[3] In the end, Professor Hill would

[3] I once distinguished, in the same sense, between 'analytic' description and 'synthetic' (*On Defining Linguistic Units*, *Transactions of the Philological Society*, 1954): but this use of two rather overburdened terms does not seem to have found much favor.

appear to demand no more. But even this is a very considerable demand to make. It gives the work its peculiarly insular character. It is responsible for its theoretic interest, and also the source of its most serious shortcomings.

The Hierarchy of Language

What I have called componential description may of course refer to any type of simpler unit—in an expansive procedure, to *any* of its lower levels, not necessarily the lowest. Professor Hill, however, has a distinct preference for going further than is necessary. "It has been our practice," he says, "to grant priority to phonological signals . . ." (pp. 356, 115). This bias certainly leads to some interesting discoveries. The phonological information we are given on every kind of English construction is both abundant and meticulous. But the *priority* which is granted to it is what raises doubts. It would appear to produce a distorted picture of the language as a whole.

The semantic or functional values of linguistic elements are pushed into the background. This creates a paradoxical situation. On the one hand, the importance of meaning and function is acknowledged as a heuristic aid; on the other, it is minimised in the subsequent 'formal' description. Having recognised and selected what are relevant elements by noting their functional characteristics, we are then required to describe them as far as we can without reference to what marks them as relevant. Such description may well fail to accomplish its purpose; it may not succeed in identifying and distinguishing all those selected elements by means of their internal features only. Functional characteristics will then be invoked as a kind of stop-gap. But—what is worse—even if a purely componential description succeeds, it can only do so at the price of distortion; it suppresses what is of primary importance.

The most important fact about the vast majority of linguistic elements is their instrumental value—their capacity of being chosen from determinate ranges and for determinate functions. To neglect this in their 'formal description' is much like trying to describe a monetary system without referring to the uses and values of notes and coins. It might be done; and componential features—like dates of issue, being more or less soiled, folded or flat, etc.—may then acquire tremendous importance. But could we be content with this kind of description? Is it not clear that, in such systems—technological, social, biological—internal 'markers' are of interest only in so far as they are *correlated* with characteristic functions? There is no reason why the essential distinctions should be made purely in terms of markers.

It may perhaps be claimed that, in proceeding from sound to sentence,

the functional value of an element, though not appearing in its description, will yet be recognised at a later stage—namely, when the element comes to be treated as a constituent of larger units. But this is not so. The larger units, in turn, are introduced without any clear notion of *their* functional significance, and are therefore incapable of imparting significance to their constituents.

A general lack of rational motivation is in fact the most unnerving feature of the present procedure. As we move along, working from smaller to larger units, we are puzzled at every step why we should go on with it. Thus, having finished with phonology, we are simply told: "The morpheme is the unit we encounter after the phoneme" (p. 89). Not a word of why we should care to make its acquaintance. Not, we are warned, because it has meaning (pp. 90, 100) nor because of its distinctive function in forming sentences (p. 92). Then why? We are not told at all. We are just invited to describe it; as far as we can, "in terms of phonemes" (p. 115). This is how we have to go on—always trying to maintain our interest in what is being described by simply at every step trusting the author that he is guiding us towards a final goal which is worth coming to. The test is "all the material which follows" (p. 93). "Beyond the morph and morpheme units, the only[!] larger units which can be observed and verified are phrases and sentences" (p. 115). Surely, a surprising statement. As it stands, it could hardly be true. And even if it were, if there were really nothing else to be "observed and verified" amongst the sequences of phonemes than just this, we should still be entitled to know why we should want to observe it. Why have them—the 'minimal words' 'inflectional paradigms', 'fixed phrases', 'free phrases', etc.—those many 'entities' between morph and sentence? That they "can be defined" and that "a phonological basis" can be laid "for definition and distinction" is surely no good reason for actually doing such things. Reasons for doing what we are asked to do are not easy to find in this book. References to meaning or function are used so sparingly or so inexplicitly, even as a heuristic aid, that the student is left unenlightened about the structural significance of the units described.

Some scholars, though working with the same general programme of analysis, have been able to avoid its worst drawbacks. They do it by being less insistent on the completeness of a componential description; allowing themselves to refer to meaning (almost invariably, when passing from phonology to morphology) or to refer to function (for example, Z. S. Harris in his *Methods in Structural Linguistics*, where every new step in the procedure is introduced by a statement of its purpose, i.e. inevitably, a statement about the distribution of the units to be described). Such deviations from the set direction of the procedure have great advantages.

Not only do they make for a more inspiring treatment of the subject—a point of some consequence in a student's text-book; what is more important, they result in a far less lopsided picture of the language.

It should be acknowledged, on the other hand, that the lopsidedness of Professor Hill's description is not of a capricious sort. It is entirely due to his consistency—consistency in pursuing a procedure once adopted "until its possibilities are exhausted." This was the promise at the beginning (p. vi), and it has been kept. All the more important is to examine the result with care; it is the test of a procedure.

Consider, first, the picture it offers us of the language as a whole. What do we make of its hierarchical structure? We have seen how weakly motivated the transitions are, from level to level. The link which connects any one with the next is tenuous in the extreme—a true bottleneck for the 'ascent'. Thus, it is only by noting junctures, especially the elusive internal 'plus-junctures' (+), that we are supposed to find our way from sequences of mere phonemes to morphs and morphemes (pp. 91ff.). It is, again, essentially by noting terminal junctures and certain stress- and pitch-features (here called 'morphemes' in a very peculiar sense, in pp. 102ff.), that we pass on to words (pp. 115ff.); and it is by seeking out certain stress-pitch-juncture constructions ('superfixes' and 'contour-linkages') that we finally reach phrases (pp. 125ff., 173), 'sentence-elements', and sentences (p. 259). What a narrow escape from being left 'grounded' with nothing to entertain us but phonemes! If we did not succeed in passing the sequences of phonemes through a process of 'junctural cutting', we should have no morphs, no morphemes; and if we did not establish those very queer stress- and pitch-morphemes and their 'constructions', our formal description of English would be without words and phrases. In this description, the *raison d'être* of units of different levels appears to consist almost entirely in various occurrences of features of stress, juncture, and pitch. Appropriately, these features are brought in with the first descriptive chapter of the book (pp. 13ff.). But for them, there could be no 'formal' linguistic hierarchy at all. They provide the narrow passage which separates and connects the successive levels. This is what it means to "grant priority to phonological signals."

In this connexion, it is curious to note that the phonology itself suffers from being granted priority. As it is made the foundation of the whole hierarchy of language, its own status is made much more precarious than it would otherwise be. Especially that triad of stress, juncture, and pitch, which is to provide the backbone of the hierarchy, appears as highly problematic. Every feature of these three types only just scrapes through a test of phonological relevance. There seem to be two reasons for this:

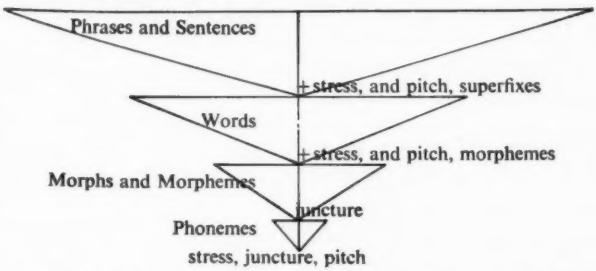
(i) that Professor Hill confines himself to a single criterion of phonological relevance—namely, contrastiveness, and (ii) that he applies this criterion to isolated stress-beats, individual pitch-levels, and single juncture-points. He frankly admits that he has experienced great difficulty in making even the native English speaker perceive such contrasts. "It is necessary to advise the reader," he says, "that if he still has trouble . . . after having worked through the possible distinctions which have been given, he should consult a competent analyst, and let the analyst sort out contrasts for him" (p. 18). A sign, surely, of having come to the very margin of normality, and therefore of linguistic significance.

Is it really necessary to get into such straits? We seem to have found two reasons for this particular predicament, and there seem to be two remedies: (i) the few phonological contrasts of stress or juncture or pitch that there are, may be accounted for more easily in terms of prosodic '*contours*' than in terms of phonemic point-features; (ii) the primary phonological relevance, even of these contours, should be found in their frequent occurrence as '*markers*' of types of syntagma, and of position-classes within such types, i.e. in their providing a regular frame for *other* units contrasting, rather than in their contracting contrasts themselves.

No one can accuse Professor Hill of ignoring prosodic patterns or contours. He mentions them at the very beginning (pp. 20–21, 26–27); and they are conspicuous and carefully treated at all levels of analysis. He describes the stress-patterns of noun-phrases (pp. 181–182) and verb-phrases (pp. 193ff.), the pitch-contours or 'linkages' of 'sentence-elements' (pp. 259–260), and of sentences within sentences (pp. 353–354). These patterns or contours are seen primarily to mark regular frames for the contrast or choice of *other* elements. Particularly interesting are some passing remarks on juncture as a marker of maximum ranges of choice (referring to work by G. A. Miller and H. L. Smith Jr., pp. 341–342). Constant concern with prosodic markers is one of the characteristic features of this book. The snag is only that the descriptions are ultimately all based on those elusive prosodic 'points' which are so hard, and so unnecessary, to track. Why, for instance, should we be asked to distinguish 4 stress-patterns in modal-verb phrases: weak-primary, tertiary-primary, secondary-primary, weak-secondary (p. 193)—and are we really sure that there occurs no tertiary secondary—if all these could be described, and easily recognised, as 'rising contours'? The answer is that if we wished to avoid 'prosodic points', and decided to treat prosodic features only as contours, and mainly as frame-markers, then we should be required, at the start, to state their grammatical scope, and thus to use the kind of information which is simply not available for a ground-level phonology.

Prosodic elements need not be phonologically elusive or marginal. As contours, they are easily perceptible; and as markers, they are centrally important. But it is impossible to introduce them as such, if we adopt an expansive and combinatorial procedure of linguistic description.[4] Once committed to such a procedure, we find ourselves in a paradoxical situation: while the greatest possible burden is laid on prosodic markers, these markers themselves can have only the most precarious phonological foundation. As elusive and utterly marginal 'phonemes', they are made to carry the whole hierarchical fabric of the language.

As we proceed from sound to sentence, the hierarchy of language presents itself as a peculiarly acrobatic structure. If the smallest 'list' of simple items is succeeded by ever larger lists of ever more complex items, the hierarchy of the successive 'lists' must appear like an inverted pyramid. And Professor Hill's hierarchy is even more fanciful. As he insists on priority for phonological signals throughout, he provides no more than a single narrow channel to connect the successive levels; nothing but stress, juncture, pitch—all the way from the apex on the phonological ground-level to the very broad base at the top. A pictorial representation of this would be a structure of a number of pyramids, all inverted, the smallest on ground-level, and ever larger ones rising above it. Not more than pinpoint contacts between them to let phonemes pass into morphemes, and morphemes into words and phrases—something that may be represented by the diagram below. (A more exact representation would contain a larger number of pyramids).



Surely a precarious structure; the slightest flaw anywhere brings it down in ruins.

[4] If treatment of prosodic features continues to be the least satisfactory part of phonological analysis, the reason seems to lie in what Professor J. R. Firth has described as "the misapplication of the principles of vowel and consonant analysis to the prosodies" (cf. "Sounds and Prosodies" in *Papers in Linguistics*, 1934–1951, pp. 121–138). It is significant that the most helpful contributions in this field have come from scholars who would not comply with the schedule of a combinatorial analysis.

It is a very different kind of hierarchy that emerges from a reductive procedure of analysis. Reductive analysis, whether it advances in a continuous fashion or not, *begins* on the broad basis of unit-utterances; and the hierarchy of language, when it finally has taken shape, is erected on the basis of this minimal abstraction from the given facts of speech. The units at every successive level emerge as a selection from the preceding; they are solidly embedded in units of lower level—every phoneme in morphs, every morph in words or phrases, and, ultimately, all in the common ground of unit-utterances. What distinguishes elements of different levels is primarily not a phonological badge they wear but different restrictions upon their 'freedom of occurrence'.

This situation is, of course, familiar from applications of Immediate Constituent Analysis—which is one of the principal techniques of a reductive procedure of linguistic description. Not surprisingly, Professor Hill has little use for it. In this respect, his treatment differs greatly from both Gleason's and Hockett's. He refers to IC-analysis only incidentally—mostly, when he suspects that it might run counter to results which he has obtained without making use of it. He then betrays a rather rigid notion of the procedure. A sentence or phrase is viewed as a strictly linear structure, and only single cuts between 'before' and 'after' are admitted. When this does not work, when the resultant description turns out to have separated what belongs together, such inadequacy is glossed over by subsequent devices of 're-assembly' and 'transcription' (pp. 279–280, 289–290, 350, 405). But these are only casual excursions from the main course of Professor Hill's procedure, where IC-analysis—like meaning—is never used as a 'primary tool'. Indeed, he tends to ignore it precisely where it might have been most helpful for supporting and expanding his own descriptions (e.g. in the treatment of noun-phrases, verb-phrases, modifying phrases, or—most conspicuously—in his discussion of 'down-grading'). Of course, if he had tried to avail himself of such support, he would have had to modify his picture of the hierarchy of language, and—consistent as always—he abstains.

It might be suggested—it certainly has been—that 'the results' may be 'the same' after all; that there may be one-to-one correspondence between the elements of an expansive and those of a reductive hierarchy. Certainly; and so there may be between the parts of two pyramids, one standing on its base and the other upside down. To climb the one is very different from climbing the other. This, no doubt, brings in extraneous facts. Upside down, we must ask, with regard to what? Necessarily, with regard to something outside the object of description. But to judge 'results', that is, to view the description of a language as a whole, *is* to view it against a

background of other facts, and to apply 'extraneous' criteria such as rational coherence, systematic density, adequacy. Thus viewed, the expansive hierarchy is a fanciful structure, rising strangely and precariously from just a mouthful of articulatory features; while the reductive structure is firmly rooted in the behaviour of people talking and listening, and emerges as a gradual analysis of their meaningful utterances. At every step, we know the value of what we are looking for—even when peering into a speaker's mouth or examining a spectrogram. The expansive hierarchy may contain the 'same' elements in the 'same' combinations; but as we meet them on our miraculous climb, they are all deprived of their values.

Synthetic Structures

The most serious result of a bias for componential description is the loss of perspective it entails. It is as if a living organism were described in great anatomic particularity, but in almost complete disregard of physiology; or if a machine were described by meticulously listing parts, without reference to how it works. The greater the profusion with which component parts, and parts of parts, are spread out before us, the harder we find it to appreciate their significance. The significance of a linguistic unit is not found in its components.

This tendency which it has to deprive us of a proper sense of orientation, is a general drawback of the expansive procedure of componential description. There are also more specific difficulties—difficulties concerning the description of *particular* structures. A componential description which is not strictly controlled by criteria of functional value is liable to give us both *more* and *less* than we want.

We are given *more* than we want, when the component 'marker' of an important distinction turns up either (a) in something that is not worth marking (as when the *-ly* of *proudly* occurs in *silly*) or (b) in something that should indeed be marked but for different reasons (as when that same *-ly* turns up in *manly* or *sickly*)—of course, a very common kind of difficulty. What is to be done about it? Simply pretend that what is marked *is* important, and always important in the same way? Would it not always be wiser to correct the bias for 'markers'? In the case just cited, Professor Hill decides between adjectives and adverbs by supplementing the componential criterion, presence of *-ly*, by distributional criteria, and also by simply listing exceptions (pp. 170–171). But he treats functional criteria always as subsidiary, and tries always to avoid them—especially in identifying the more basic units. This can be troublesome. What, for example,

shall we think of a procedure which analyses *potato* into 5 morphemes: {ˇ} {pə} {+ˊˇ} {teytə} {231#}? This is Professor Hill's analysis (p. 120). And he is only consistent in proposing it. When juncture is treated as the primary and regular signal of a morphemic boundary, we must be prepared to find ourselves flooded with synthetic elements of this kind; which—as Professor Hill admits—"are not required by analysis of larger entities" (p. 92). They are, in fact, not required by anything or anybody.

This unwanted influx is something Professor Hill would make light of. He does not seem to mind if important functional distinctions are blurred, so long as they do not prevent him from going on to 'higher levels' (p. 92). He can put up with two segmental morphs *dam* and *nation* appearing in *damnation*, or with *at all* being analysed as /ə/ + /təhl/ (p. 92); for, queer though they are, these elements are excused on the ground that they come to join up in a proper way, when we meet them later as parts of larger structures. But this is strange argument. To allow it would be to abandon all check on the adequacy of analysis on any level; there would only be one requirement—namely, that an element should be part of a sentence. It is as if in describing a furnished room we allowed ourselves to pick out, say, a floorboard supporting two wooden poles as a 'piece of furniture', seeing that "nothing on the higher levels prevents such cutting." Is it not true, after all, that in the ultimate description of the room, the two floorboard-poles will join up with two others, and with a table-top, and with other floor-boards? And when this happens, will not our procedure have been tested and declared adequate "by its final results?" Is there nothing wrong with it then? Shall we say that it just uses the term 'piece of furniture' in a novel sense? Shall we say of Professor Hill's use of the terms 'morph' and 'morpheme' that it is just unusual? I think, our objections must be more serious. For the new use sponges on the old. The whole justification of proceeding with a morphological analysis—Professor Hill's as much as any other—derives from the old sense, which is here perverted. The reason why we describe a room in terms of 'pieces of furniture' is that these are the more stable units with which we find we have a certain freedom of operating within the room, as we have with 'morphs' and 'morphemes' within a sentence. If we try to distinguish morphs primarily by the presence of 'junctures', just as if we tried to mark off pieces of furniture by the presence of cracks and empty spaces, we find ourselves inundated with 'pieces' we have no use for—though they are all pieces, ultimately, within our universe of discourse. Such elements do not modify morphological analysis; they make it pointless.

To 'build up' linguistic units that are not wanted is, then, a serious flaw. Equally serious, of course, would be a failure to supply linguistic units

that *are* wanted. Here, however, Professor Hill is often prepared to correct the deficiency of a componential description. We have seen that he makes use of contrastive distribution and of syntagmatic relations, whenever the progress of his analysis requires it. "Sounds are grouped into phonemes primarily in terms of their distribution, not in terms of their physical similarities and differences" (p. 47). The same is said of grouping morphs into morphemes (p. 94). Distributional parallels are also required for establishing morph-boundaries by 'second cuts', i.e. "where there is no /+/ [juncture] to mark them" (pp. 94, 97–98). And an appeal to 'syntactic characteristics' becomes ever more frequent in the second half of the book. Even so, distributional criteria are kept to a minimum, and always to a subsidiary role. If it is at all possible, they are stopped short of developing into semantic criteria. This, again, tends to keep them too lax, and unwanted 'elements' rush in, or distinctions which are needed are not made. I cannot see, for instance, how Professor Hill's criteria for 'second' cuts can prevent us from analysing, say, *stable* into the 'morphemes' {s} and {table} or *stone* into {s} and {tone}, in much the same way in which *tables* are analysed into {table} and {z} and *tones* into {tone} and {z}; or why, on his criteria, we should object to analysing *battle*, *cattle*, *handle* into {bæt} + {l}, {kæt} + {l}, {hænd} + {l} (cf. pp. 96–97). "Phonemic similarity," says Professor Hill, "is a necessary criterion; meaning is not." In order to recognise a morph (such as *table*, *tone*, *cat*) as recurring, he asks only for phonemic similarity and for similar grammatical distribution (such as we find in the utterances *I saw the cat*/*the cats*/*the cattle*). Hence *foxglove* is said to contain allomorphs of the same morphemes as occur, say, in *foxhunt* or *silkglove* (p. 100).

Generally, distributional criteria will be kept to a minimum, with the result (i) that Professor Hill's English has surprisingly few homonyms, and (ii) that his description does not seem to have any place for 'idioms'. This causes grammatical difficulties at times. When idioms get mixed up with genuine constructions, the latter, and their elements, may not be easy to classify. The following is a characteristic case: *He ran up a hill* and *He ran up a bill* are found, in many dialects, to have the same characteristics of juncture and stress—namely, those which Professor Hill regards as a marker of prepositional constructions /əpə+/. The second sentence, *He ran up a bill*, can also occur with what he calls adverbial stress /ə́pə+/. What is it then, preposition or adverb? "If we assume ... that stress difference is always distinctive, then we are forced to say that either adverb or preposition can occur in this particular sentence." This is what Professor Hill does say, "since otherwise we are ultimately making an appeal to meaning" (pp. 284ff.).

Such blind reliance on markers raises doubt. If distributional criteria were used more extensively, the whole question would not arise. The conundrum would be solved by finding that the idiomatic *up* of *ran up a bill* is neither preposition nor adverb; it is not to be identified with either the preposition of *up a hill* or the adverb of, say, *the balloon went up*. Similar difficulties recur quite frequently. Is it satisfactory to describe *to* as a 'preposition' in *He wants to go* (pp. 245, 250), and as an 'adverb' in, say, *When Mary came to, I comforted her* (p. 255), or to describe *daggers* as a noun-complement to *looked* in *John looked daggers at Mary* (the pattern being that of *John saw daggers*, p. 326–327)? These seem to be cases of distinctions needed but not made—distinctions between homonyms, or between idioms and constructions. But we must acknowledge that no really satisfactory way of dealing with either seems to be available at present; and, as always, Professor Hill prefers formal surgery to informal patching up.

There is perhaps some justification for judging a procedure by its excesses. The occasions for excesses are different for different procedures. Professor Hill's favorite occasions are those cases of asymmetry, where functional distinctions or classifications, though themselves unavoidable, fail to correlate with componential. Here, the search for components is to be seen at its highest pitch of ingenuity; it may even go so far as to conjure up what is inaudible to human ears.

All of this is pseudo-analysis: the attempt to provide diacritical components where there are none. Familiar amongst such devices is the introduction of *additive* zero-elements (as distinct from that legitimate element zero which marks the contrastive *omission* of some overt element or elements). For example, instead of being content to state that a word like *sheep* shares its grammatical functions partly with words like *ox* and partly with words like *oxen*, we are invited to locate this difference of functions in a difference of indiscernible components, i.e. to contrast *sheep* with *sheep* as 'zero absent' and 'zero present'. Again, instead of merely stating that /bɔyz/ shares the functions of *child's*, *children*, *children's*, we are supposed to distinguish by components /bɔy −z, bɔyz −, bɔyz ø/ (pp. 138ff.). If these diacritical symbols were recognised as purely functional marks—mere class-indices—all would be well. But they are not. They are made out to mark components. 'Zero' is described as an 'allomorph', and even said to contrast "only with its absence" (p. 143). It is never clear, how far we may go in fabricating such ghostly components. The contrasts *am ÷ are*, *was ÷ were*, or the numerous contrasts of the kind *drank ÷ drunk*, *drove ÷ driven*, are not made the basis of zeros in other paradigms (although such zeros would satisfy the 'alternation-rule' (p. 455)); while contrasts

like *drink ÷ drank* are supposed to justify a distinction between *cut* and *cut* ø, or again, (I) *drink* ÷ (he) *drinks* is supposed to justify a distinction between (*I*) *can* and (*he*) *can* ø (p. 161). *Who* in *Who are they?* is supposed to have "a zero allomorph of the plural suffix" in concord with the verb (p. 263). The great danger of the procedure is that it can be applied very much more widely than even Professor Hill has cared to apply it. If zero were used here as index of a purely functional classification, it would mark the componential *defectiveness* of a paradigm, and thus indicate important asymmetries of the language. But if we use it as a component element, we pretend to supply the defectiveness, and we disguise the asymmetry which we ought to mark.

The most spectacular and most ingenious of these pseudo-analyses of Professor Hill's is perhaps his treatment of the personal pronouns (pp. 145ff.). Again there seem to be good reasons for establishing the 'paradigms' themselves, i.e. for collecting the forms in parallel sets of maximally four distinctions (*I, me, my, mine*; *he, him, his*; *etc.*). The distinctions and parallels can be justified on functional grounds. Professor Hill admits that this might be all that "a teacher of English as a second language" requires. But he thinks that "from the point of view of the analyst," the forms of the different sets should be shown to *contain* those parallel distinctions, and contain them as inflectional endings. Now, of the seven sets, only one (*they*) shows anything like a full inflection with a more or less constant base—namely, /ðey, ðe-m, ðe-r, ðe-r-z/. It takes great ingenuity to extend the same kind of inflection to the other six sets. In three of them, a second basis is required (*I, me*; *we, us*; *she, her*); in many cases, we have syncretisms (*you, his, her, it, its*); in some, zero-allomorphs of suffixes (*me, us, his*); and, in two, even portmanteaux: "{miy} plus {-r} equals {may}," and {may} plus *-n*, which is taken to be an allomorph of {-z}, brings {may-n} into line with /ðe-r-z/. When this 'practical' part of the analysis, as it is called, is completed, the given 23 units appear as built up in complicated ways out of exactly 23 morphs. Nor is even this the end. Further analysis—"complete, nonpractical and abstract"—of the basis (*I, we, they*, etc.) would seem to add another 11 morphs (pp. 148ff.). (*It*, for instance, would seem to "contain" three morphs: transposed "initial" {t}, transposed {i}, and zero-allomorph of {y}.) In the final analysis, then, 23 units appear to be analysed into 34 morphs, of which none has any use except that of analysing the 23.

This kind of description is a long way off the general reductive purpose of analysis; and Professor Hill tries to justify it on other grounds. His analysis, he wishes to claim, reveals internal formal parallels, where otherwise we should have only syntactic similarities (pp. 148ff.). It will be

acknowledged that an analysis of different forms may be worth while for the sole reason that it shows the forms to be of similar structure. Even if we do not succeed in reducing the number of elements, something will be gained if we succeed in discovering mutually corresponding components in the different forms. The trouble with Professor Hill's procedure is only that it could hardly fail. His operations of "consistent cutting and statement" are such that they can extract different components from identical forms, and similar components from utterly different forms. (I cannot even see, why he thinks he would have to abandon his analysis of pronouns, if his material contained the pronominal forms *thee* or *hit* (p. 150). He could surely succeed in 'normalising' these two, as he has succeeded with others.) The search for components has almost succeeded, here, in freeing itself from the control of given facts. The elements we obtain are so elusive that no speaker of the language could be credited with the power of operating with them, and no hearer with the power of responding to them. Would it not be better to regard such elements, and their supposed correspondence, as illusory, precisely because they are "non-practical"? They seem to add nothing to that similarity of syntactic function which they were invoked to support. Our concern with language is, after all, concern with a pattern of human behaviour; what no one can react to, cannot be part of it.

The magic of pseudo-analysis has a certain fascination for some linguists; however irritating it may be for others. Of the latter, many might be tempted to put this book away before they have finished with it. For them, the section on pronouns might well be the last straw. It would be a pity. For when they have reached chapter 9 on the inflection of nouns, pronouns, and verbs, the most useful part of the book is just to come. This part—not surprisingly—is its syntax (chapters 11 to 20). It is here, in the exploration of larger structures, that a componential description of English can begin to be really informative—even if an excessive bias for this type of description can still be seen to be a drawback.

We have found (pp. 255, 261 above) that, in entering upon his analysis of phrases and sentences, Professor Hill begins to make extensive use of distributional criteria. It is only at this stage that an expansive procedure *can* use them more extensively. Though the basic elements have been defined already and defined mainly by their components, *classes* of these elements will now be defined frequently by "syntactic characteristics." There is a variety of such characteristics, corresponding to the variety of analytic operations by which they are obtained. Professor Hill has a distinct preference for one type—namely, 'order-classes'. His favourite procedure, here, is to set them up for the maximum of a particular type of

construction—a maximum noun-phrase, for instance, such as *all the ten fine old stone houses* (pp. 175ff.), or for a maximum modifying phrase (pp. 232ff.). Shorter constructions are then derived from the maximum model. (Indeed an analogous technique has already been used in Professor Hill's phonology—in what is perhaps its most interesting section, entitled 'phonotactics', which deals almost exclusively with consonantal clusters (pp. 69ff.).) In analysing verb-phrases, Professor Hill finds it more convenient to begin with three simple types (chap. 12) and to obtain larger phrases by expanding the basic (chap. 13); but again, he will give close attention to limitations of order (pp. 222–223).

What we find in these chapters on English syntax, will, it seems to me, be appreciated by every one. It remains valuable, even if we are not content to rest in it. We may well wish our syntactical investigations to be less confined to considerations of linear order; we may wish them to be kept open to a greater variety of analytic techniques. With Professor Hill, it happens only once—in a rather desperate case (p. 401)—that he has recourse to 'diagnostic frames' of the kind used so extensively by Harris and Fries. Again, 'selection' (concord, government), supremely important for Bloomfield and for European linguists, is used only twice (pp. 222, 260). Operations of transformation (such as have lately been stressed by Chomsky) are not explicitly used at all. (They would have helped formally to distinguish what for Professor Hill are utterances of the same type, e.g. *This seems a big price* from *This brings a big price*, p. 293; similarly, p. 323.) We have seen that Immediate Constituent Analysis has been assigned a very minor and subordinate role. Had it been applied more freely, a more significant hierarchical organisation would have been found in what appears here often as purely serial order.

A definitive English syntax seems to be beyond anybody's reach at present. Professor Hill's study, however, is an important step towards it. Even the most controversial of his interpretations provide an impulse for fruitful development. I only regret that it is impossible within the limits of a review—even of a review-article—to illustrate by specific examples how much there is to be learned from Professor Hill's English syntax. Though the topics themselves are traditional, the treatment of these topics cannot be discussed by just holding it against the tradition and noting a few deviations. We are dealing here, in other words, with an original work—original in an important way.

This may well be a book that had to be written, an experiment (as Professor Hill would call it) that had to be made. It will furnish students of the English language with enough to occupy them for a long time to come. For there is nothing in it that does not deserve their interest and scrutiny,

and perhaps nothing, either, that would strike them as final. We shall have to pass beyond this book. But there can be no doubt that it gives us valuable, perhaps indispensable, help for doing just that.

Department of General Linguistics
University of Manchester

Reviews

A. I. SMIRNICKIJ, *Leksikologija anglijskogo jazyka.* 260 pp. Moskva: Izdatel'stvo Literatury na Inostrannyx Jazykax, 1956.

Reviewed by D. S. WORTH

This posthumously compiled volume makes available the frequently original and stimulating opinions on English lexicology held by the late Professor A. I. Smirnickij of Moscow. It was put together from lecture notes, articles, and unpublished papers by a group of Smirnickij's students, and occasionally betrays its scissors-and-paste origin (e.g. a passage on the derivational function of the paradigm occurs on pp. 24–25 and again almost word-for-word on p. 77). It presents a generally coherent and systematic discussion of the major problems of English and general lexicology, but remains within the framework of a series of provocative Fragestellungen rather than attempt even tentative solutions. A brief introduction delimiting the aims and content of lexicology (pp. 5–11) is followed by a theoretical background section entitled "Basic Problems of the Word" (pp. 12–47) and five chapters devoted to word structure and derivation (pp. 48–137), the interrelations of phonetics and semantics (pp. 138–173), classification of the lexicon (pp. 174–202), phraseological units (pp. 203–230), and etymology (pp. 231–255). In what follows, I shall try to outline some of Smirnickij's more interesting viewpoints, dwelling on them in detail only where questions of principle seem to be raised.

The introduction insists that the lexicon of any given language must be described as a coherent whole, and complains that all lexicological works to date have devoted relatively little space to the mass of everyday words that make up the great bulk of the lexicon. While not denying the importance of the diachronic viewpoint (inherent in any synchronic slice listing productive and nonproductive categories), Smirnickij dismisses historical lexicology as not really lexicology at all, since it does not treat the lexicon as a system but only points to its origins. This reviewer does not see any reason, however—at least in principle—why the lexicon of a given language could not be described diachronically as the historical evolution of a system and the subsystems thereof, just as historical phonology and morphology are presumably describable in systematic terms; the basic

and still unsolved problem is to discover the systematic framework in terms of which the lexicon can be described.

The section called "Basic Problems of the Word" opens with a cursory bow toward the Marxist definition of language (a collectivity of lexicon and grammar, in which lexical units are the building materials and grammatical structure the organizing force), and then passes to a definition of the "linguistic units" of which language consists. Smirnickij would accord status as "linguistic units" only to those elements of language in which there is a marked correspondence of form and meaning. Thus, for him only words and morphemes are "linguistic," since they and only they have meaning of their own. Phonemes (and presumably distinctive features, although this level of analysis was apparently not yet being used in Moscow when these notes were set down), which have no meaning of their own, can at most be considered units out of which "linguistic units" are constructed. It is wrong to say that morphemes consist of phonemes, because only the external form of a morpheme is divisible into smaller discrete units, whereas its meaning, an integral part of its structure, cannot be so divided. This distinction between "linguistic" and non-linguistic units is basically a terminological one (since, after all, everyone knows that morphemes have meanings of their own and phonemes do not), but Smirnickij's terms seem less than felicitous, since it might well be argued that any type of unit, at any level, which functions in a demonstrably systematic manner in the total system of systems that is language, might well be called a linguistic unit.

In the other direction, Smirnickij states that word-combinations and sentences cannot be considered "linguistic units," because they arise only during actual discourse and, unlike morphemes and words, do not have an independent existence outside of the concrete instances of speech. Smirnickij seems to go off the deep Marxist end in saying that if all concrete word-combinations were "linguistic units," then everything said and thought would be part of linguistic structure, instead of simply being expressed by this structure. Here it can be replied that the word-combination itself exists as a structural type, independent of the particular lexical items which fill it in any given utterance (e.g. Russian has such a combination type "past passive participle plus substantive in the instrumental case," the most frequent meaning of which is "action and actor thereof" and tokens of which are *prinesen studentami* 'brought by the students,' *pročitannaja nami* 'read by us,' etc.). Smirnickij compares word-combinations to such numeral combinations as 197, 1543, etc., which are themselves not part of the numeral system. This comparison is less striking when we remember that unlike numbers, word-combinations exist in

strictly limited types and are strongly differentiated in kind: whereas it might be hard to describe the structural difference between, say, 98 and 87, there is no difficulty in describing the difference between, say, possessive substantive plus substantive on the one hand and adverb plus verb on the other. This neglect of or lack of interest in the structural importance of units above the word level is in my opinion the major weakness of this book. It is of course only natural that a scholar whose principal interest lay in lexical description and dictionary-writing (the best Russian–English dictionary was produced under his guidance) should concentrate on the word itself. Still, since we know so little about syntax, and next to nothing about the relation of syntactic structure to the lexical system (if there is such, which fact has also still to be demonstrated), it seems hasty to assume a priori that no units larger than the word can play any functional role within the lexical system. This is too much like throwing out the bath water before washing the baby.[1]

The remainder of this section discusses the word from the lexical and grammatical points of view, and gives the criteria upon which Smirnickij distinguishes between varieties (lexical or grammatical) of a single word on the one hand and separate words with common (lexical or grammatical) elements on the other. Since the word is characterized by its paradigm, a difference in paradigms which places two forms in two different parts of speech (e.g. *love* in the set *love—loves—loved—loving*[2] vs. *love* in the set *love—loves—love's—loves'*) makes two separate words of these two forms. If paradigmatic distinctions do not place two forms in two separate parts of speech, then these forms are to be considered separate words only if some paradigmatic difference between them corresponds to some difference in their meanings: the formal difference between *hanged* and *hung* makes separate words of *hang* (by the neck until dead) and *hang* (a picture on the wall), but in the absence of any difference in meaning *learnt* and *learned* are only formal varieties of the single word *learn*, just as *shade* (of a tree) and *shade* (of green) are only lexical varieties of the one word *shade*. This distinction, although not uninteresting, has two weaknesses. First, a "difference in meaning" is a hard thing to define objectively, and there are surely many degrees of difference: who is to determine in a given case whether two words have different meanings or only nuances of the same

[1] Later in the book (p. 157) Smirnickij does admit that "lexico-semantic" variants of a word "find their expression" in differences of syntactic structure and combinability with other words, but he does not draw the apparently obvious conclusion that the correlation of form and meaning obtains at the syntactic as well as at the morphologic level.

[2] Smirnickij uses spelling forms rather than morphophonemic formulae in all his descriptions.

and still unsolved problem is to discover the systematic framework in terms of which the lexicon can be described.

The section called "Basic Problems of the Word" opens with a cursory bow toward the Marxist definition of language (a collectivity of lexicon and grammar, in which lexical units are the building materials and grammatical structure the organizing force), and then passes to a definition of the "linguistic units" of which language consists. Smirnickij would accord status as "linguistic units" only to those elements of language in which there is a marked correspondence of form and meaning. Thus, for him only words and morphemes are "linguistic," since they and only they have meaning of their own. Phonemes (and presumably distinctive features, although this level of analysis was apparently not yet being used in Moscow when these notes were set down), which have no meaning of their own, can at most be considered units out of which "linguistic units" are constructed. It is wrong to say that morphemes consist of phonemes, because only the external form of a morpheme is divisible into smaller discrete units, whereas its meaning, an integral part of its structure, cannot be so divided. This distinction between "linguistic" and non-linguistic units is basically a terminological one (since, after all, everyone knows that morphemes have meanings of their own and phonemes do not), but Smirnickij's terms seem less than felicitous, since it might well be argued that any type of unit, at any level, which functions in a demonstrably systematic manner in the total system of systems that is language, might well be called a linguistic unit.

In the other direction, Smirnickij states that word-combinations and sentences cannot be considered "linguistic units," because they arise only during actual discourse and, unlike morphemes and words, do not have an independent existence outside of the concrete instances of speech. Smirnickij seems to go off the deep Marxist end in saying that if all concrete word-combinations were "linguistic units," then everything said and thought would be part of linguistic structure, instead of simply being expressed by this structure. Here it can be replied that the word-combination itself exists as a structural type, independent of the particular lexical items which fill it in any given utterance (e.g. Russian has such a combination type "past passive participle plus substantive in the instrumental case," the most frequent meaning of which is "action and actor thereof" and tokens of which are *prinesen studentami* 'brought by the students,' *pročitannaja nami* 'read by us,' etc.). Smirnickij compares word-combinations to such numeral combinations as 197, 1543, etc., which are themselves not part of the numeral system. This comparison is less striking when we remember that unlike numbers, word-combinations exist in

strictly limited types and are strongly differentiated in kind: whereas it might be hard to describe the structural difference between, say, 98 and 87, there is no difficulty in describing the difference between, say, possessive substantive plus substantive on the one hand and adverb plus verb on the other. This neglect of or lack of interest in the structural importance of units above the word level is in my opinion the major weakness of this book. It is of course only natural that a scholar whose principal interest lay in lexical description and dictionary-writing (the best Russian–English dictionary was produced under his guidance) should concentrate on the word itself. Still, since we know so little about syntax, and next to nothing about the relation of syntactic structure to the lexical system (if there is such, which fact has also still to be demonstrated), it seems hasty to assume a priori that no units larger than the word can play any functional role within the lexical system. This is too much like throwing out the bath water before washing the baby.[1]

The remainder of this section discusses the word from the lexical and grammatical points of view, and gives the criteria upon which Smirnickij distinguishes between varieties (lexical or grammatical) of a single word on the one hand and separate words with common (lexical or grammatical) elements on the other. Since the word is characterized by its paradigm, a difference in paradigms which places two forms in two different parts of speech (e.g. *love* in the set *love—loves—loved—loving*[2] vs. *love* in the set *love—loves—love's—loves'*) makes two separate words of these two forms. If paradigmatic distinctions do not place two forms in two separate parts of speech, then these forms are to be considered separate words only if some paradigmatic difference between them corresponds to some difference in their meanings: the formal difference between *hanged* and *hung* makes separate words of *hang* (by the neck until dead) and *hang* (a picture on the wall), but in the absence of any difference in meaning *learnt* and *learned* are only formal varieties of the single word *learn*, just as *shade* (of a tree) and *shade* (of green) are only lexical varieties of the one word *shade*. This distinction, although not uninteresting, has two weaknesses. First, a "difference in meaning" is a hard thing to define objectively, and there are surely many degrees of difference: who is to determine in a given case whether two words have different meanings or only nuances of the same

[1] Later in the book (p. 157) Smirnickij does admit that "lexico-semantic" variants of a word "find their expression" in differences of syntactic structure and combinability with other words, but he does not draw the apparently obvious conclusion that the correlation of form and meaning obtains at the syntactic as well as at the morphologic level.

[2] Smirnickij uses spelling forms rather than morphophonemic formulae in all his descriptions.

meaning? Secondly, this distinction may be a static oversimplification of a dynamic process: it might well be maintained (although hard to prove) that language tends to maintain a functional correspondence of form and meaning, which means that differences in meaning tend to express themselves by developing differences in form (this would account for the reformation of *housewife* opposed to *hussy*) and differences in form tend to be preserved by the development of new, or the attribution to them of old differences in meaning (as with *hanged*, *hung*). To maintain an obsessive distinction at all points between "one word" and "two words" is to overlook the real nature of this continual process of adjustment.

Chapter III discusses in some detail the morphemic structure of English and the various procedures of derivation and composition evident in the English lexicon. Much of this material is standard, and has little relation to the systematic description of the lexicon which the introduction leads one to expect. Smirnickij pays particular attention to the interrelations of the various parts of speech, and to "conversion" (paradigmatic derivation, as *love* verb vs. *love* substantive). Here as elsewhere he appears to exaggerate the importance of certain polar oppositions and to underestimate the importance of the underlying features upon which these oppositions themselves are based. On pp. 72–77, for example, Smirnickij discusses at length whether *love* substantive and *love* verb are one word or two. However, if we agree with Smirnickij that there is a strong opposition of lexical and grammatical morphemes in English (and there are those who would not), and that the word is characterized by its paradigm as well as its meaning, then we can say that *love* verb is opposed to *love* substantive by an opposition of grammatical features (letting V = verb paradigm and S = substantive paradigm, *love* + *V*::*love* + *S*), just as *hate* verb is opposed to *hate* substantive. Then, *love* verb is opposed to *hate* verb and *love* substantive to *hate* substantive by an opposition of lexical features (their respective "meanings," however the term is defined; these can be represented by the dictionary forms *love*::*hate*). In addition, we might say that *love* substantive and *lover*, *hate* and *hater* are opposed by a derivational feature (both lexical and grammatical in nature). We would then obtain a network of oppositions

love + V	love + S	love + r + S
hate + V	hate + S	hate + r + S

which, *mutatis mutandis*, is rather like a network of phonemic oppositions based on the presence or absence of certain distinctive features, e.g.

/n/	/d/	/t/
/m/	/b/	/p/.

Of course, there are limits to the productivity of such inter-level analogies; the basic point is merely that one must constantly bear in mind all the units of which a given system is made up and all the features which unite or oppose each unit within the system to each other unit. Smirnickij has taken only a few tentative steps in this direction.

Chapter IV, "Phonetic and Semantic Characterization of the Modern English Word," begins with some correct but rather superficial phonetic remarks concerning An- and Auslaut phenomena, consonant clusters, etc., and ends with an overly long and at times scornful criticism of Skeat's etymological dictionary. In between, there is an interesting but occasionally one-sided discussion of problems of meaning, including polysemy and homonymy. As might have been expected, Smirnickij censures the *fizikalistskaja škola L. Bljumfilda* for wanting to eliminate meaning from linguistics. For him, meaning is neither the object to which a word refers, nor the relation between this object and the word itself (he does not attribute these definitions to any source, nor does he mention any serious theories of meaning), but is "a reflection of an object, occurrence or relation in the (human) consciousness . . . (which) enters the structure of a word as its so-called internal side" (p. 152). Although Smirnickij is by and large a devout believer in the duality of lexicon and grammar, he does note that no meaning can ever be entirely lexical, at least not in an inflected language (including English), since the very fact of belonging to a certain part of speech necessarily adds a certain grammatical element to the basic semantic content of the word. "Consequently, the grammatical classification of words, expressed in their distribution as parts of speech, is from the lexicological point of view a particular semantic classification—a classification according to the most generalized and abstract component part of their meaning" (pp. 152–53).

In discussing polysemy, Smirnickij wishes to distinguish between (1) different uses (pointing to different signata) of one and the same lexical variant of a word and (2) two different lexical variants of a word. His example, however, is so chosen as to provide a perfect case of the investigator's native language structure and the thought patterns connected therewith being superimposed upon a second language under investigation and then as it were "discovered" within this second language. Smirnickij's explanation speaks for itself on this point, and in addition provides a quaint commentary on the inadequacy of the translation process in general (essential explanations have been added in square brackets);

> "Thus, for example, the word *oil* [Russian *maslo*] can signify a series of different oils, with the exception of cow's (for which the word *butter* is used) [Russian *maslo* also = butter]. However, it does not follow from this that in denoting various oils, the

> word *oil* has a different meaning each time: in all cases it has exactly the same meaning, namely 'oil (any kind except cow's)' . . . It is a different matter, however, when the word *oil* denotes 'oil' [Russian *neft'*=crude or fuel oil]. Here one has in the foreground not the similarity according to oiliness of oil [*neft'*] with various kinds of oil [*maslo*], but a particular quality of oil [*neft'*]—combustibility. And here the word *oil* in this meaning is correlated with words denoting various types of fuel, such as *coal* 'coal' [Russian *ugol'*], *wood* 'wood' [*derevo*], *fuel* 'fuel' [*toplivo*], etc. This makes it possible to separate out within the word *oil* two meanings, and by this token two lexico-semantic variants of this word: (1) the variant *oil* 'oil' [*maslo*] and (2) the variant *oil* 'oil' [*neft'*]" (p. 157).

The facts of the matter seem to be that the great variety of oily substances in the real world is simply divided at different points and by different criteria in Russian and English. The Russian distinction of *maslo*::*neft'* classifies oils according to origin, secondarily according to function (*maslo* can be solid or liquid, used for eating, cooking, or lubricating, but is in all cases derived by some process from animal or vegetable substances; *neft'* has various industrial uses, but in all cases flows out of a hole in the ground), whereas the English distinction of *butter*::*oil* seems to be that between a solid table spread (cf. *apple butter*, which contains nothing oily at all) and all oily liquids regardless of origin or function (cf. *cooking oil*, *machine oil*, *oil wells*, etc.).

Smirnickij distinguishes among three types of relation among lexico-semantic variants of a word: metaphoric (*deep* river::*deep* sorrow), metonymic (container::contained, as big *kettle*::boiling *kettle*), and synecdochic (*man* in general::*man* vs. woman). Instead of pursuing this division into more interesting detail, however, he spends the remaining ten pages of this chapter on a partially justified but disproportionately long and detailed critique of the treatment of homonyms by Skeat.

In Chapter V Smirnickij names two principal types of semantic classification. The first or "logical" classification divides words first into parts of speech, then, within each part of speech, into broad semantic groups (it is not made clear just how one is to define semantic groups), then into smaller subgroups, and so on. English substantives, for example, can be divided into words denoting animate beings, concrete inanimate objects, processes and actions, qualities abstracted from objects, etc. (one wonders where the "etc." really leads). The group denoting animate beings can be divided into groups such as that of nomina agentis, itself divisible into a group of "accidental" actors (*bystander*, *onlooker*) and a group denoting specialists (*teacher*). No formal criteria for separating out these various groups are given, nor is any sizable group of words classified according to this schema. Since the author himself notes that the most variegated "logical" classifications are possible for the same group of words within the same language

(depending on the hierarchal order in which classificatory criteria are applied), one is left with the discomforting impression that neither Smirnickij nor his pupils have applied this schema to a large number of randomly-chosen words. Only then could one judge whether or not there is anything systematic about this "logical" classification.

The second or "thematic" classification of words is based on the existence of such "themes" as time, space, number, size, position, change, etc.[3] Smirnickij feels that such a classification will make clear the different semantic structures of different languages (e.g. the dichotomy of *come*::*go*, *kommen*::*gehen*, *venir*::*aller* vs. the single Russian *itti*). Unfortunately, his examples do not do much to accomplish this. The fifteen-odd pages devoted to a "sample analysis of a group of words denoting 'direction of motion' by thematic classification," do little more than give an inventory of the morphological devices by which motion can be expressed. Smirnickij specifies that any discussion of motion must take into account five aspects: (1) the moving object itself, whether (in the non-grammatical sense) "subject" (*the horse runs*) or "object" (*the horse is led out*); (2) the type of action (*run, walk, crawl*); (3) the general direction of the action (whether "toward" or "away"); (4) the object or place in relation to which the motion occurs; (5) the specific nature of the relation between motion and object or place (*into, out of, past*). The remaining discussion ignores the first three points altogether (thus eliminating much of the lexical material dealing with motion and concentrating on the grammatical or semi-grammatical), and gives a list of the various parts of speech by which points (4) and (5) are expressed. For example, (5) the specific nature of the relation between motion and place can be expressed by special adverbs (walk *across*, take *off*), prepositions (walk *across* the street, aller *à* Paris), verbal prefixes (*en*tomb, *ex*port), verbal roots (*monter*, *sortir*) and by various compounds of the above, such as the specifically Russian combination of preverb and semantically identical preposition (usually historically identical as well), e.g. *vyjti iz* komnaty 'go out out of the room' or *v*"exat' *v* gorod 'go in into the city'. Although this section contains a number of interesting observations, particularly on the various types of pronominality, it does not even resemble the sort of fairly inclusive if not yet very detailed semantic classification of the English vocabulary that one had been led to expect. As a matter of fact, I could not even discover the "themes" according to which the classification was presumably made. The rest of the

[3] It would be interesting to see a semantic classification that did not start with such "themes," all of which can be lumped together under the general head of "quantifiers" (cf. my notes on this in *International Journal of Slavic Linguistics and Poetics* I/II [1959], 117f.).

chapter consists of traditional remarks about stylistic differentiation, in particular the role of slang.

Chapter VI divides phraseological units into two major types, according to whether they contain a single major semantic element (lexical morpheme) or more than one such element. The first, "single-peak" type contains such varieties as verb plus adverb (*give up*, *bring up*, *make out*), verb *be* plus adjective (*be surprised*, *be tired*), and preposition plus noun (*for good*, *by heart*, *in time*). In all cases, the meaning of the phraseological unit is more than (=different from) the sum of the meanings of its parts. The second, "two-peaked" or "multipeaked" type has at least two major semantic units. Here belong combinations of adjective plus substantive (*best man*, *first night*, *black art*), substantive plus substantive (*son-in-law*, *man-of-war*), and verb plus substantive (*catch cold*, *fall in love*), plus what Smirnickij terms "repetitions" (*povtory*) and what Malkiel (*Lingua* VIII, 113f.) has called "irreversible binomials" (*now or never*, *with might and main*). In addition, Smirnickij distinguishes traditional combinations (*clench the teeth*, *shrug the shoulders*) and true idioms (*take the bull by the horns*). A few complete sentences also function as such ready-made units (*how do you do*, *thank you very much*). This chapter, while rather poor in examples, provides an interesting and adequate description of phraseological units in English.

Chapter VII, "Etymological Characteristics of the Modern English Lexicon," begins with a generalized discussion of the evolution of form and meaning, the borrowing process, relations of words in cognate languages, etc., and proceeds to a brief description of Scandinavian and French loanwords in English prior to the fourteenth century. The book ends, as it were in mid-air, with the statement that the historical evolution of the English vocabulary will be studied in the author's forthcoming *History of the English Language*.

To sum up: this book is an uneven mixture of the stimulating and the trite. It is peppered throughout with acute observations and original suggestions. Had the author lived to complete it himself, it would undoubtedly be a far more finished product. As it is, however, it does not provide, nor even come close to providing, a systematic description of the English lexicon. Putting this book down, one is regretfully aware that a real structural lexicology is just as far away as ever.

University of California
Los Angeles 24, California

JACOB M. LANDAU, *A Word Count of Modern Arabic Prose*. XX+453 pp. (+20 pp. in Arabic and 13 pp. in Hebrew). New York: American Council of Learned Societies, in cooperation with the School of Education, Hebrew University, Jerusalem, 1959.

Reviewed by MAJED F. SA'ID

Teachers of modern literary Arabic, particularly those interested in writing primers, graded readers and the like will welcome the appearance of this volume designed to serve as a research aid.

The volume consists of an Introduction (in three versions: English, Arabic and Hebrew) and two Parts. In the introduction the author surveys the available literature on the subject and notes some of the shortcomings of existing materials for the teaching of modern literary Arabic. Here he underscores the lack of balanced vocabulary selection. He then proceeds to outline the method employed in his count, the guiding principles for which are those laid down by Brill in 1940.[1] Landau devotes a good portion of his introduction to a description and evaluation of Brill's list. This list, restricted to the vocabulary of Arabic daily journalism, is incorporated in this volume side by side with Landau's list for modern Arabic prose based on sixty books written by modern Egyptian authors. These books are a fair representation of modern Arabic prose.

The number of running words counted by Landau was 136,089. This is the same as the number of words counted by Brill. From the count emerged 11,284 specific words. This is almost double the number of specific words (5,981) in Brill's list. In the combined version, there are 12,400 specific words. This number is obviously less than the total number of specific words in the two separate lists, the reason being that many words are shared by the two lists.

Part I of the volume is the alphabetical list. Here all the words listed by Brill and by Landau are arranged according to the alphabetical order of the Arabic roots, and are classified either under "L" (Literature) denoting the frequency in Landau's count, or under "N" (Newspapers) denoting the frequency in Brill's list. A third column gives the total of the two frequencies. The entries in this part are printed side by side with a transcription[2] in Latin characters and are provided with English glosses. The

[1] Moshe Brill in collaboration with D. Neustadt and P. Schusser, *The Basic Word List of the Arabic Daily Newspaper*, Jerusalem: The Hebrew University Press Association, 1940.

[2] The transcription system is a modified form of that used by Hans Wehr in his *Arabisches Wörterbuch für die Schriftsprache der Gegenwart*, Leipzig: Harrassowitz, 1952.

addition of transcription and glosses is a commendable feature of the work.

Part II is the frequency list. Here the specific words which appear in the alphabetical list are arranged according to their frequency in the 272,178 running words of the combined list; the total is divided between the "L" and the "N". Also included are numbers denoting the range of a word in the total of 12,400 specific words. These enable one to locate easily the 10, 25, 50, 75, 100, etc. most frequent words in the combined list, a remarkable feature of great service to textbook writers.

The following results of the count are worthwhile noting: the first 25 specific words in the frequency list represent 23.70% of the total number of running words; the first 500 represent 58.53%, and the first 1,000 represent 70.02%. Thus if a student masters the 1,000 most frequent specific words in the Brill–Landau combined list, he would (presumably) be able to comprehend—as far as vocabulary is concerned—700 words out of 1,000 in the daily newspaper or in modern Arabic prose. This is a discovery of no small importance and will at once point to the practical value of this study. However, this reviewer in no way intends to give the impression of placing an undue emphasis on vocabulary selection in the preparation of good teaching textbooks. Other factors, such as the morphological and syntactic patterns of the language, are equally if not more important. The main point to bring out, however, is the potential usefulness of word counts in the grading of language textbooks. This becomes very important in the case of Arabic teaching materials, the majority of which suffer from a lack of balanced selection and graded build-up of basic vocabulary; add to this the fact that Arabic boasts of a rich and extensive vocabulary, a fact hardly ever encouraging or consoling to the beginning student.

The dependence of Landau on Brill's rules for counting calls for a comment. Landau follows generally the technique and rules established by Brill twenty years ago. This might be justified by the desire of the author and the publisher to facilitate comparison between the two counts. True, the author questions Brill's sampling technique (p. ix) and tries to remedy some of its limitations; for example, he attempts, with a large measure of success, to cover a cross-section of 20th century Arabic prose, whereas most of Brill's newspaper material came from sources that appeared during a period of less than two years. Brill's approach is open to criticism, hence one wishes that Landau had set up new rules for counting, even at the expense of foregoing comparison between his count and that of Brill's. One would have liked to see verbs listed with the prepositions that characteristically accompany them, especially since Landau has already deviated —with justification—from Brill's rules for the listing of prepositions. Although this is not a major departure, yet in a careful statistical work it is

important enough to raise some doubts about any serious attempt at exact comparison. Another undesirable carry-over from Brill is the discrepancy between the treatment of the active and passive participles on the one hand and verbal nouns on the other. Whereas verbal nouns are listed separately as a class regardless of their function, active and passive participles are sometimes counted as substantives and other times as verbs. Verbal nouns, too, can function as verbs and as substantives. Would it not be preferable if the active and passive participles were also listed separately each as a class by itself? This of course brings up the very important question of whether to use function or form-class as the criteria for counting. Brill, a pioneer in Arabic word counts, cannot be blamed for not facing this problem as early as the publication of his list—1940. Might not the decision to use one criterion or the other affect the reliability of Landau's word count?

In a spot check of words in transcription, a fair number of typographical errors were noticed. Also, the Arabic introduction suffers from a stilted style and a number of errors of usage that are sometimes incomprehensible without recourse to the English version.

On the whole, the present work is a painstaking and commendable job and will remain a useful index for Arabic word frequencies. In addition to its use in controlling the vocabulary of primers and graded readers, it will be of interest to people engaged in literacy programs, language statistics and general lexicographical work.

Department of Oriental Studies
Princeton University
Princeton, New Jersey

W. KAPER, *Kindersprachforschung mit Hilfe des Kindes*. xxiii + 244 pp. Groningen: J. B. Wolters, 1959.

Reviewed by JEAN BERKO

The subheading of this book, *Einige Erscheinungen der Spracherwerbung erläutert im Lichte des vom Kinde gezeigten Interesses für Sprachliches*, indicates the framework within which the author has attempted to present his abundant observations on the linguistic activity of his two sons. The text is in German and the examples are in the children's native language, Dutch. The notes on which the work is based cover a period of somewhat

over six years, at the end of which time the elder son was about eight years old and the younger was about six. The book is divided into 37 sections, plus an ample bibliography, subject and name indices, and very helpful summaries in German, French, and English.

An attempt has been made by the author to organize his material around those aspects of language that children themselves comment on and draw attention to. In the first respect the book is properly a metalinguistic work —a study of language about language. In so far as it also treats children's unusual formations, it oversteps these limits, since the child in producing his own analogical forms is not talking about language, and it is the observer and not the child who calls attention to these forms. This fact does not, of course, reduce the value of the observations, and since many of them are of interest, they may induce researchers to investigate certain phenomena, either through experimentation or further observation.

In particular, the author feels that a Jakobsonian principle of maximal contrast can be applied to all phases of language learning, including writing. In one of his examples, he notes that his son at the age of 4.6.1, when trying to write a word with two *e*'s, wrote one with two *9*'s instead, and himself commented, "*Papa, als je twee negens achter elkaar zet, dan lijkt het ee*" (p. 188). Both the action and the comment indicate that this child could not distinguish between an *e* and what is essentially the same character printed backwards, a *9*. The ability to contrast "backward" and "forward" visual forms is obviously necessary for learning to write—*p* and *q* have become proverbial because of the problem they present. Investigators might tell us just what other contrasts the child must learn to make, which he learns first, which are the last and/or most difficult to acquire before his reading and writing reach the level of the adults around him.

The introduction and first two sections of the book discuss adult influence on the particular types of expressions children use, and emphasize the individual differences observed in the author's two sons: one son willingly and openly practiced new words and expressions, while the other, despite all efforts to the contrary, would only practice in secret and when he thought no one could overhear him. It is also noted that children's conscious interest in language and spontaneous comments about it may be greatly affected by the amount of adult attention directed at these activities. Sections 3–8 deal with children's critical comments about their own and other's choice of words and grammatical and syntactic forms, with the last section illustrating various sorts of play with the forms of address. Section 9 notes that well after a child has been taught and knows formal grammatical rules, he continues to produce forms at variance with those rules and from his own inner patterning. Parts 10–16 discuss the discovery of

partial resemblances in words, opposites, and various contributing factors in the production of new words.

Then, in the following three sections, a rather startling conclusion is reached, namely that children follow no system in forming derivations, compounds, and inflexions. This is surprising because the examples cited can be used to show just the opposite. For instance, in section 19, titled "*Auch Flexionsformen weisen keine konsequent durchgeführte Systematik auf*" (p. 72) the author explains that although his seven year old son knew that when asking for one six cent streetcar ticket he should say *één van ses*, when asking for two such tickets he said *twee van sessen* instead of *twee van ses*. This is by direct analogy with *één mes* 'one knife', *twee messen*, and many other similar plural forms in Dutch, and can hardly be an example of no system.

The sections on derivation and compounding lend themselves to the same sort of criticism. Numerous children's words are examined, and they do, indeed fall down in a careful linguistic analysis; *unsafeable* by analogy with *unbreakable* (p. 65) leaves something to be desired (this example was quoted from another researcher), but it does not really prove that child productions are characterized by lack of a consistent system. If a child makes up a perfectly good word, which he does every time he produces a regular plural he has never heard before, no one notices it. The words he produces that are not in common use may or may not be in accordance with a formal linguistic analysis of what should take place. But the failure to be entirely consistent does not seem to be the special province of children, as witness forms like *amputee* and the *Crocker-Anglo Bank* in English. On the other hand, I felt that the children's forms cited displayed a remarkable—if limited—consistency. It is not at all clear that the child who at the age of 5.9.0 made up the word *unsafeable* instead of *unsafe* had enough words in her vocabulary ending in *-able* for anyone to expect her to recognize the separate meaning of this morpheme, while the use of the prefix *un-* to mean *not* is in itself a pretty good example of consistency and rather sophisticated generalization.

Sections 21–31 comment on the questions posed by children, the construction of sentences, and the various phenomena of self correction. The final sections of the book are concerned with written language, and contain some hypotheses on how children learn to read, with further observations on the effect of reading on speaking.

As Dr. Kaper himself comments, these notes are intended to stimulate further research and not to present final answers. Many of the points made have already been elaborated by Jakobson, Guillaume, Leopold, and other observers of child language. A clearer picture of the actual linguistic

development of the author's two sons might be obtained if the text had a few tables showing the age at which various phenomena occurred, but nonetheless many ideas for future research may be obtained from this book.

Institute for Defense Analysis
Princeton, New Jersey

ARMAS SALONEN und PAAVO SIRO, *Studien zur neusumerischen Syntax*. Helsinki, 1958.

Compte-rendu par RAYMOND-RIEC JESTIN

Ce petit ouvrage grammatical a pour base les études récentes d'Adam Falkenstein sur les textes juridiques sumériens et on s'y efforce à l'aide d'exemples pris dans ces textes d'établir des règles de syntaxe avec une assurance que les faits ne me paraissent pas toujours purvoir justifier. Le point de vue des auteurs est d'ailleurs délibérément unilatéral: ils ne font pas porter leurs recherches ailleurs que dans le domaine qu'ils examinent et ne se réfèrent qu'aux travaux d'A. Falkenstein.

Dans leur premier chapitre (pp. 12 s.s.), on trouve une affirmation selon laquelle les éléments infixes dans le complexe verbal *-n-* et *-b-* renvoient au sujet lorsqu'il s'agit du présent et à l'objet lorsqu'il s'agit du prétérit. Il convient de remarquer tout d'abord qu'il est très regrettable d'une façon générale de voir se maintenir ces termes de «présent» ou «présent-futur» et «prétérit», encore communément employés par les sumérologues quand ils ne correspondent pas à la conception sumérienne du temps. Le *-e* caractéristique du «présent-futur» peut apparaître avec un «prétérit» et inversement le «présent-futur» peut, comme souvent le «prétérit», être dénué de cet indice. Il ne s'agit pas de formes aberrantes, elles s'expliquent très simplement si l'on considère le *-e* en question comme un élément traduisant une idée d'«actualisé».

En ce qui concerne, ensuite, la réalité des faits invoqués, on est bien obligé de constater la validité de la base sur laquelle ils s'appuient et qui consiste en une citation d'un passage du *Poème d'En-mer-kar*, interprété contrairement à Samuel N. Kramer et à moi-même, de façon telle que la justification recherchée paraisse établie. Cette nouvelle traduction ne s'impose cependant en aucune manière.

Il faut bien reconnaître que cette règle est applicable en tel ou tel cas, inapplicable en tel ou tel autre, car en vérité ces éléments *-n-* et *-b-* figurent ou non dans des expressions identiques et le point de vue des auteurs ne s'adapte qu'à des aspects possibles de son mécanisme lequel peut être interprété de plusieurs façons dont aucune n'a réussi encore a s'imposer.

Il est toujours loisible, en effet, de dire: Ici le rédacteur a trouvé bon de mettre l'accent sur la relation verbe-sujet, grâce à l'emploi d'un infixe approprié, là il a préféré ne pas le faire. En d'autres cas encore, c'est sur l'objet qu'on a jugé bon de mettre l'accent ou de ne pas le mettre non plus. Car tous les exemples envisagés entrent nécessairement dans l'une de ces catégories. Et il faut bien tenir compte de l'absence des éléments considérés au même titre que de leur présence. Il n'y a donc pas à s'étonner de la multiplicité des théories émises au sujet de ces infixes sans d'ailleurs que les traductions des textes en soient affectées; on saisit le sens de la proposition sumérienne, on ne perçoit pas toujours le caractère véritable de certaines catégories grammaticales et trop fréquemment apparait chez les sumérologues la tendance à retrouver en celles-ci celles qui leur sont familières ou connues, alors que manifestement certaines sont irréductibles à quelque concept actuel que ce soit. C'est le cas pour les préfixes; qu'en est-il pour les éléments *-b-* et *-n-*? On y a vu tour-à-tour des facteurs de représentation du sujet ou de l'objet, des indicateurs du mode causatif des prolongements phonético-emphatiques, des marques de la conjugaison objective. La tentative de spécialisation plus grande faite dans l'ouvrage considéré s'ajoute à ces théories, sans entraîner plus de certitude. Les auteurs reconnaissent d'ailleurs qu'il rencontrent «einige Ausnahmen» ce qui est beaucoup trop modeste.

Dans la seconde partie de l'ouvrage qui traite de l'emploi des postpositions, on trouve une nouvelle «règle», symétrique de la précédente, selon laquelle la postposition renvoie au sujet lorsque le verbe de la proposition est intransitif et à l'objet lorsque ce verbe est transitif.

Une première remarque, encore, parallèle également à celle que j'ai faite à propos de la première règle, s'applique aux verbes eux-mêmes dont le caractère transitif ou intransitif reste à démontrer, autrement, tout au moins, que par des paradigmes arbitraires qu'aucun examen critique n'appuye. Plusieurs fois mise en doute, la réalité des formes transitives et d'intransitive du verbe sumérien a été récemment formellement niée et sur le terrain du raisonnement, par Viktor Christian: «Das zweigliedrige Konjugationsschema lässt aber auch keine Möglichkeit für die Annahme zu, das Sumerische habe die Unterscheidung von transitiv, einerseits, intransitiv-passiv andererseits gekannt, eine Unterscheidung, die nach

allem, was wir vom Sumerischen wissen, dieser Sprache von Haus aus offenbar völlig fremd war» (*Beiträge*, p. 25).

Et, d'autre part, toujours également dans l'esprit des commentaires précédents, on trouve des postpositions là où on en avait besoin et pas ailleurs, sans autre considération: ainsi, si dans un exemple comme N. *nam-dam-še ba-an-tu*$_{12}$ (*S*) 'N. pour épouse a pris', on trouve la postposition *-še* et *si*, dans un autre comme *á arad-da U. L. -ra in-na- ab-rúg-rúg*: 'le loyer de l'esclave U. à L. a remplacé (ersetzen)', on n'en trouve pas qui corresponde, c'est-à-dire qui affecte le régime, c'est que la postposition s'imposait dans un cas et pas du tout dans l'autre. Que le verbe soit considéré comme transitif ou intransitif, il en est évidemment de même partout: ainsi par example dans: *sig*$_4$*urú-ab-ba-šè im-ši-sìr-sìr-e-dè-es* 'vers les murs de la cité de leurs pères ils crient' où l'on voit employés à la fois la postposition *-šè* et l'infixe verbal correspondant *-ši-*.

Une dernière partie est consacrée aux verbes composés. On trouve ici une observation très juste, si elle n'est pas nouvelle, sur l'impossibilité d'établir une différence entre verbes qui sont composés et verbes qui ne le sont pas. Il convient aussi de considérer en plus la nature des éléments dont il s'agit, le substantif et la racine verbale. Ainsi *šu . . . ti* 'main' + 'prendre' = 'prendre', un tel verbe composé ayant en définitive le même sens que la racine verbale seule. Ceci s'explique par l'analogie de la fonction d'un substantif comme *šu* avec celle des déterminatifs qui indiquent la matière dont est fait un objet p.ex. *ĝiš al* ('bois') 'houe'. On y apporte simplement l'dée d'organe (*ĝiš gin-gin* 'membres'). De même *šu* caractérise les actions faites par la main. Cependant certains substantifs de verbes composés ne possèdent pas cette nature et comme ils apportent une notion aussi ou presque aussi importante que celle que traduit la racine verbale, le complexe verbal ainsi constitué n'équivaut plus à la seule racine; le régime est alors indirect: *šag̃*$_4$ 'coeur' + *kúš* 'opinion, sentiment' = 'méditer, tenir ou prendre conseil.'

Ecole des Hautes Etudes, Sorbonne
Paris

Manual of Phonetics, Edited by L. KAISER, Secretary of the Permanent Council for the International Congresses of Phonetic Sciences. xv + 460 pp. Amsterdam: North-Holland Publishing Company, 1957.

Reviewed by MARTIN KLOSTER JENSEN

Phonetics is not easily defined or described by immanent means. Phonetic units, e.g. "sounds," are motor, physical, or perceptual phenomena which call for a definition in physiological, acoustic, or psychological terms, respectively. This is why there are at least three basic sciences in phonetics, and why branches of these sciences come to be part and parcel of it. There is, in fact, no specific phonetic method. This explains the very composite character of Elementary Phonetics.

The aim of this *Manual* is "to bring the very best special knowledge in a general and readable form" (p. ix). This of course means a compromise between "special" and "general," the more so as "the Manual is not intended for a special public," though "supposed to be read by engineers and linguists, by surgeons and psychologists, by teachers and artists" (p. ix).

The wide field of Phonetics is surveyed in six Parts: A. The History of Phonetics, B. Basal Sciences of Phonetics, C. Phonetic Research, D. Phonetics in Its Relation to Linguistics, E. Phonetics in Its Relation to Other Sciences, and F. Phonetics and the Origin of Speech.

For reasons already mentioned, it is most practical to treat these parts separately, with occasional comparing remarks.

Part A. The History of Phonetics

Chapter 1. *Earlier History of Phonetics*, by G. Panconcelli-Calzia, pp. 3–17. There can have been no question about assigning the task of writing this chapter to the grand old man in Experimental Phonetics, the author of the *Bibliographia Phonetica*, which appeared in *Vox* almost up to World War II, and of several books on the history of Phonetics. Panconcelli-Calzia successfully supports the view for which he has been known for the last fifty years, that Phonetics is not a daughter of Linguistics, nor any *ancilla demissa*. "The facts are rather (1) that phonetics takes its origin from the magic-mystical and partly religious mental world of mankind in a time, moreover, in which there were neither grammarians nor linguists, and (2) that it owes its development chiefly to scientists and physicians" (p. 17).

Chapter 2. *Technical Development of Phonetics*, by F. TRENDELENBURG, pp. 18–28. Readers who miss illustrations of apparatuses mentioned, should remember that the stress is laid on showing what can be registered more than on giving details about instruments. After a survey of early methods in sound analysis and the production of artificial speech, the author deals with the mechanical recording of speech, the evaluation of such recordings, and recording of sound by electrical means, among them

refined types of playback devices. Three pages are devoted to automatic electrical methods of sound analysis and electrical production of synthetic speech, comprising a wide range of instruments for acoustic analysis of the speech signal. However, no mention is made of "pattern playback."

Part B. Basal Sciences of Phonetics

Chapter 3. *Morphology and Physiology of the Speech Organs*, by G. E. ARNOLD, pp. 31–64. The six sections of this article deal with respiration, the larynx, the voice, the pharynx, "oral structures," and the nose, respectively. We shall confine our remarks to three of the sections.

Section 3. The voice. As the presence of partial tones in the primary laryngeal sound (p. 47) does not seem to have been frequently or thoroughly discussed in phonetic literature, it would be desirable to have at least some reference for an explanation. The work referred to (Joos, *Acoustic Phonetics*) hardly accounts for more than the filtering performed by supraglottal articulation (Joos, p. 37–42). In connection with vocal ranges (p. 48), space may have forbidden a discussion of the measurement of *tessitura* (personal medium voice range), in terms of recurrential chronaxie, on the m. sternocleidomastoideus, reported by Husson (*Bull. Acad. Nat. de Médecine*, nos. 1 et 2, 1955, p. 25).

Section 4. The Pharynx. The role of this cavity as a resonator for the voice is well explained. In a half of a page (p. 51) the reader will find fundamental knowledge about coupling and damping.

Section 5. Oral structures, contains a rather rigorous statement which will not easily pass unnoticed: "A complete and correctly occluding set of teeth is indispensable for faultless articulation" (p. 52). Good remarks are made about degrees of elevation of the palate for different speech sounds (p. 55). The roentgenogram facing p. 56 is not particularly clear: the reader may be very little familiar with X-ray pictures. Marking of the velum would have been helpful (so also on p. 176). The physiologist duly stresses that "human speech sounds are more than the sum of their underlying motions" (p. 58).

Chapter 4. *Auditory Considerations*, by N. R. FRENCH, pp. 65–93. This contribution is concerned with the relation between auditory sensations and "the physical patterns of signals." Within the auditory system physical disturbances take place, the knowledge of which "is still in its infancy" (p. 65), so, for the present, it is out of the question to "deduce the brain patterns from the known patterns of acoustic signals and then interpret them in terms of auditory sensations" (Section 1).

Section 2. Physical aspects of auditory system, intends to "bring out

some of the ways in which signals are altered and distorted in their progress through the system" (p. 69). It is stated that an individual hair cell in the Organ of Corti "is made available for use by a wider range of frequencies at higher intensities," whereby the auditory system is able to "overcome some of the inherent limitations of nerve fibers" (p. 71) which react *in groups* on stimuli, and can therefore transport to the brain ranges of intensity and frequency much wider than those which can be reported by one single nerve fiber.

Section 3. Characteristics of hearing, deals with pressure thresholds, loudness measurements, masking, pitch, effects of intensity level and of duration on pitch, and with the number of discriminable steps in frequency and intensity.

Section 4. Auditory sensations and their interpretation, is largely concerned with the limitation on the use of auditory sensations in the analysis of signals (p. 82) and the types of information which the organism perceives, namely: level, form, and duration. It is pointed out that "in general, the total sensation is not an aggregate of the sensations produced by the separate components, but rather a new sensation in which the identity of the components seems to be lost" (p. 88). This new sensation, "closely akin to (a) single, coherent sensation" (p. 91) illustrates the difference between "brain patterns" and "the known patterns of acoustic signals."

This brilliant chapter should be preliminary reading to all students of acoustic phonetics.

Chapter 5. *Phonetics from the Viewpoint of Psychology*, by A. GEMELLI and J. W. BLACK, pp. 94–117. The two sections deal with "phonetics for the listener" and "phonetics for the speaker."

Section 1. An outline is made of "the relative intelligibility of phonetic units" (pp. 95–96), and "the probability of occurrence of the phonetic unit" (pp. 102–107). The question of minimal duration of perceptible speech sounds, and of maximum distortion of recognizable sound spectrum ("temporal and spectral thresholds") are discussed (pp. 107–109). Results of experiments with pattern playback and "reordered speech" are reported to show the role of phonetic environment for the interpretation of speech sounds; the conclusion is drawn that the listener's reception of the phonetic signal is partly conditioned by knowledge of the probability of occurrence and of the listener's motivation towards reception of the signal.

Section 2, about the role of the speaker in speech psychology, discusses the effects of delayed feedback, and of sound pressure level in one's own voice as perceived in feedback. It also relates of experiments to yield a scientific basis to the common experience that the degree of precision in

articulation depends on the pattern of the stimulus (unconscious imitation of the interlocutor).

The authors, in 22 pages, have done much to show that the science of phonetics is closely related to psychology (if not part of it), both as regards perception and production of speech.

Chapter 6. *Statistical Methods in Phonetics*, by EDWIN B. NEWMAN, pp. 118–126. There probably never was a phonetician who did not feel the need for statistics. Unfortunately, statisticians who come to the help of the phonetician often mistake phonemics for phonetics. In the present case, the material used for illustration is one of syllable counts. For better orientation about statistic procedures, the author could have chosen measurements, say, of duration and intensity of physical features of speech sounds. Anyhow, the article does give a simple formula for finding standard deviation (p. 119), it brings noteworthy remarks on "random choice" (p. 122) and "adequacy of sampling" (p. 122), and reminds the reader that "any generalization from statistical evidence is valid for the material, and *only* for the material, from which the sample has been drawn" (p. 123).

Readers interested in Zwirner's Phonometry will miss a brief outline of pertinent procedures, but the author is of course not to blame on this point.

Part C. Phonetic Research

It can be no accident that this part is written exclusively by Americans, who, more than any other group, seem to carry on intensive work in experimental phonetics. The author of chapters 7–10 is G. E. PETERSON.

Chapter 7. *The Phonetic Evaluation of Signals*, pp. 129–138, starts with the important statement that linguistic conditioning is fundamental to decision in phonemic judgment about a signal. The author examines the different objectives in the experimental study of speech (pp. 130–131), then analyses the communicative process and mechanism of speech, which brings him to a discussion of the complex acoustic waves and the relationship of physiological to acoustical structure. As a substitute for as yet not available methods of neural observation, recourse must be had to other levels of speech, namely articulation movements, sound waves, and auditory perception. For vowels, acoustical descriptions are preferred (p. 136), but we are reminded that "the phonetic evaluation of a speech signal is very much a matter for operational definition" (p. 138).

Chapter 8. *Breath Stream Dynamics*, pp. 139–148, deals with major problems of "interaction between physiological movements and the

contained air within the vocal tract" (p. 239). The interest is first concentrated on the sublaryngeal pressure variations during speech and silent breathing. Other sections are concerned with pressure fluctuations at syllable rates (pp. 144–147), and with the relation between movement of the body wall and increased chest pressure (p. 146). From Stetson's kymographic recordings (*Motor Phonetics*, p. 33), the author finds it "very probable that a certain amount of the chest pulse pattern is controlled by back pressures from the larynx and articulators, as would be expected of such an acoustical system" (p. 147). He does not doubt that "pressure fluctuations within the thorax occur at syllable rates," and that they are supplied by the muscles of the thorax and abdomen, but he finds further research needed "to determine the extent of these effects at normal conversational rates" (p. 147). The last section discusses the observable audio-frequency vibrations of "various body structures" during speech, of which those captured by thoracal contact microphones seem to be the most important in communication. One single sentence in this chapter is not clear to the reviewer: "Various vocal noises can be produced, however, with closed glottis or on inhalation" (p. 144). We can follow the author only if "vocal noises" here means "speech noises," such as voiceless stops or fricatives.

Chapter 9. *Laryngeal Vibrations*, pp. 149–155, outlines "the complex mechanical system" of laryngeal vibrations, summarizes stroboscopic findings, relates of mechanical models of the human larynx, and consecrates a part to laryngeal photography, with some observations about the mode of vibration (open phase and closure). In the last section, Laryngeal theory, the author states that "much further experimental study and theoretical analysis is required for a complete description of laryngeal operation" (p. 155).

Chapter 10. *Articulation*, pp. 156–165. On the basis of Pike's *Phonetics* (1943) and the *Preliminaries* (1952), Peterson proposes "to discuss briefly the various physiological mechanisms of speech sound production" (p. 156). The "elementary sounds" produced by the human vocal mechanism are—with due reservations—"separated into their major classes": *plosive*, *vibratory*, and *fricative*, or, acoustically, *pulse*, *periodic*, and *random* (pp. 156, 157). The description of these sounds is made on the physiological and acoustical levels. On p. 160 we find a table of "elementary plosive formations." The reviewer was at first slightly confused until he discovered that "horizontal rows" (p. 160) means the same as "the articulations listed down the side of the table" (p. 159). The page should be turned by 90° for easy reading. It contains no reference to "velopharyngeal" (Pike: velic) closure, though in fig. 12, Ab, the velum is said

to be closed for the "voiced bilabial implosive followed by *a*." This would be an instance of velo-pharyngeal closure. Though we are not going to question the justification of it, it should also be noted that vocal cords in oscillation are here considered to form a closure, as far as we can see, even a "driving closure." In fig. 12, Aa, the "laterally released lingua alveolar implosive on mouth air" would be more of a "click" in API terminology, since there is no indication that the "driving closure" is not "lingua velar."

The author has seen no reason, or need, to give more than a sketch of the articulation of speech sounds, partly for lack of space, partly because the details of physiological shapes in sound production are generally familiar to students of phonetics. This may well be so, but the acoustics of speech (movement of air in the resonating cavities) seems to call for some analysis, and can hardly be considered as well known to students of phonetics, to say nothing of the wider circle of readers. The editor may again be to blame for the fact that this *Manual* nowhere discusses the creation of vowel or consonant formants or the generation of noise in the vocal tract.

The author accounts for the fact that frictional sounds "are normally produced by means of respiratory movements" by referring to "the quantity of air requested to sustain the sounds" (p. 163). A no less satisfactory explanation would perhaps be that the construction of the larynx seems to necessitate production of all normal voiced sounds on exhaled breath, so that all unvoiced sounds, which as a whole are far less frequent than the voiced ones, are also articulated on expiratory air.

For good reasons, the complex question of concomitant articulations could be hardly more than hinted at. An illustration of sound combinations is given in fig. 12, D, but one asks why a nonsense passage has been constructed to illustrate a phenomenon which is a *sine qua non* in speech.

In less than 40 pages, Peterson has very succinctly outlined the essentials of Experimental Phonetics.

Chapter 11. *The Application of Roentgenography in the Study of Speech*, by J. D. SUBTELNY, S. PRUZANSKY, and J. SUBTELNY, pp. 166–179.

If we realize that the tongue positions for back vowels suggest a line slanting towards the front from cardinal *a* to *u*, this is, as far as the knowledge of the reviewer goes, due to X-ray research. The discovery may not seem too important at first, but it should be remembered that it breaks the parallelism between the formant charts and the sketches of tongue positions. The authors of this contribution give a survey of the use of roentgenography in speech physiology with some orientation about procedures for analysis. Of special interest is the section on Laminagraphy. In the opinion of the authors, X-ray motion pictures will undoubtedly bring out new and significant information in the future (p. 178).

This part of the book ends with two chapters written by H. L. BARNEY and H. K. DUNN.

Chapter 12. *Speech Analysis*, pp. 180–201. Due place is given to the Sound Spectrograph and to Visible Speech patterns, including so called dynamic spectrograms, and translators (where no delay occurs in the transformation from speech to spectrograms). In fig. 3, the reader will find the word "you" instead of the expected "cree", cf. p. 183. The "correlatograph" is described on pp. 185–186. It may be that the reader is to blame if he does not understand what additional knowledge he can derive from correlatograms, but it may turn out that the authors should have been more communicative on this point.—Various methods of mechanical speech recognizers ("voice type writers") are mentioned.

Section 2 deals with speech power and spectra, and brings tables showing average power of speech sounds and average fundamental frequencies. When ɔ is called "the strongest vowel," and when it is said that "the first formant is the strongest" (p. 191), then "strong" is not necessarily synonymous with "loud." This has a certain importance for the reading of fig. 6, average spectrum of conversational English speech, male and female. The reader may miss some comments on the level/loudness relation. One may also mark intelligibility directly on the basis of intensity level. This procedure has been described under the heading "Effects of Distortion on Understanding of Speech" (pp. 193–197). Section 3 (pp. 197–201) discusses the general assumptions in the statistical investigation of speech sounds.

Chapter 13. *Speech Synthesis*, pp. 202–212. A difference is made between *synthetic* speech (mere putting together of component parts [Voder, Vocoder, Pattern Playback]) and *simulated* speech (electrical analogs, artificial larynx). The section on Artificial Speech in Communication seems to contain a promise of much success for some important projects.

Part D. Phonetics in Its Relation to Linguistics

a. General Part

Chapter 14. *Phonology in Relation to Phonetics*, by R. JAKOBSON and M. HALLE, pp. 215–251. The *Preliminaries to Speech Analysis* mark an invention which, for its idea, can be compared to the art of printing: the combination of comparatively few units to constitute a large number of higher units. Letters make up "words," distinctive features combine to make "phonemes." The binaristic view of distinctive features offers more answers to *why*'s in phonetics than any other method of analysis known to the reviewer. Yet it may not be what it sets out to be, the final establishment and arrangement of the elements in the signal code of speech. If a

speech sound is heard, it is still possible to imagine a direct association with a total pattern being in turn the realization of a unit in the chain of segments where it is capable of marking, or differentiating, or obscuring meaning, or where it displays certain distributional properties. In other words, if a phoneme sequence is sounded and strikes the ear, is "auditioned," it is possible to interpret the complex of sounds perceived. Under such circumstances the conclusion that binary choices take place until the signal is finally analyzed and all possibilities of phonemic interpretation other than one have been eliminated is not necessarily the nearest one. We seem to receive messages more by evocation of concepts than by elimination of what is not in the message. The Saussurean and the Prague viewpoint in this matter sprang out from the idea of differentiation; the evoking property of the sign, or even of the phoneme, was not so much stressed. As elimination is nearer to differentiation than to evocation, it is understandable that in the *Preliminaries* elimination and not evocation became the leading idea about the reception of a message.

Binarism, choice of 'nought-or-one', very much bases its principles on observations in child language. Still, it does not support the natural point of view that the child acquires the language of its surroundings by *combining concepts with complete messages*, not by analysing the message received. No definite borderline can be drawn between the language of a child who understands common messages, and adult language, so it cannot be seen how grown-ups can be said to have left the stage where they received a message by association. Only in exceptional cases, says Householder (*For Roman Jakobson*, p. 236), when both contextual and morphemic matching fail, "do we finally resort to phonemic matching." Though speech can be analysed into phonemes, and phonemes into distinctive features, the question is: is the differentiating function of such features important enough to play a part in the acquisition of spoken language?

One of the assumptions of distinctive feature analysis is stated on p. 218: ". . . one must recur to the fullest, optimal code at the command of the given speakers." Such full, optimal codes, however, are not acquired in the same way as we learn a foreign language from a teacher, who will correct our pronunciation till he is satisfied. Our pronunciation of our mother tongues is probably as clear as it has to be and as muddled as we can allow it to be. If this is so, there is room for certain differences between the "fullest, optimal code" of one speaker and that of another, since such "full codes" are varieties which the speaker has perhaps never heard, but only constructed as an exaggeration of the pronunciation which he imitates, in other words, a product of his own analysis of language. So far,

most of the acoustic analyses we have seen are made from the speech of phoneticians or of other speech-minded subjects. It remains to be seen whether speech is analysable in distinctive features when the code is less clear, though fully understandable, i.e. as spoken solely with the object of communication, as we believe it to have been by most people in most time. The statement on p. 220 that circumstances *may cause* redundancies to substitute for distinctive features, seems hardly to do justice to facts in everyday language. The authors find that "the auxiliary role of redundancies must not be underestimated" (p. 220). There can be no doubt that such substitution for distinctive features is the rule.

Except for its polemic tone, slightly out of place in a *Manual*, the survey of different views on the phoneme in relation to sound is good reading. A distinction is made between "inner, immanent" approach, which locates the distinctive features and their bundles within the speech sounds (Bloomfield, the authors), and the "outer" approaches, which "divorce phonemes from concrete sounds" (p. 219). It is interesting to note that, while to Hjelmslev *immanent analysis of language elements* does not include acoustic or physiological investigation of speech sounds (substance), for Jakobson and Halle *immanent phonemic analysis* means extraction of relevant features from the substance.

Though some linguists, most recently Malmberg (1955–1957), have been tempted to search for substance criteria for syllable division, the majority of phoneticians seem to have lost faith in the syllable as a phonetic unit. They welcome attempts to explain its existence on the phonological level. It has been maintained that different languages follow different rules in considering certain sounds as syllabic in certain positions. A more stringent formulation is that found on p. 226, which starts from distributional criteria, saying, "The phonemic structure of the syllable is determined by a set of rules and any sequence is based on *the regular recurrence of this constructive model*" (italics mine). E. Haugen has proposed that the syllable be defined as *the smallest unit of recurrent phoneme sequences*. (*For Roman Jakobson*, p. 216.) But Jakobson and Halle also make attempts to describe the motor and acoustic correlates to what they call the "phonemic syllable" (p. 227).

The formulation of the "universal invariance in the phonemic patterning of language" has been constantly occupying the author of the famous *Kindersprache, Aphasie und allgemeine Lautgesetze* (1942). The latest formulation is found in the *Manual*, p. 232: "The inherent distinctive features which have so far been discovered in the languages of the world . . . amount to twelve oppositions, out of which each language makes its own selection." These twelve oppositions are then presented in the form of nine

"sonority features" and three "tonality features," named partly from spectrographic phenomena, partly from physiological events (justification for terminology, p. 239, middle). The oppositions are characterized in both acoustical and physiological terms (pp. 233–235). Other distinctive features are "prosodic," displayed only by phonemes forming the "crest of syllable" (pp. 227–229).

According to the authors, phonemes are "bundles" ("admissible combinations") of "concurrent (distinctive) features" (p. 217). The code includes phonemes, and therefore it contains combinations of features. The "*coding rules*," however, by which "we must be guided . . . to find out what motor, acoustic and perceptual elements of sounds are utilized in a given language" (p. 236), is a term which is not clear to the reviewer. In the "operational hierarchy of levels" in the decoding of the message, the perceptual one is ranked first, then comes the aural (?), acoustical, and motor levels (p. 237). It will be agreed that "the sporadic help of lip-reading" (p. 237) is the only justification for considering the motor aspect at all in this connection, but it must not be forgotten that lip-reading is important in acquiring a language because we have an audio-motor reflex, or semi-reflex, giving us experience on the motor level while receiving a message. On the other hand, it can hardly be seen how or when the acoustical level comes in in decoding a message, if not in laboratories. It is trivial to remark that it is only by means of perception (auditory, visual) that we can utilize the "motor" and "acoustic" levels.

The "nuclear syllable" (p. 239), the one with which child language "ordinarily begins," is said to be the utterance "usually described as /*pa*/." In the *Kindersprache* . . . Jakobson repeatedly (e.g. p. 12, note 1; p. 37) stresses the importance of distinguishing between what is purely phonetic and what is phonemic. The notation /*pa*/ leaves no doubt that *p* and *a* are to be considered as phonemes. Now, if they are phonemes, only a "perceptual, acoustic or physiological" analysis can decide what phonemes they are. If a baby is capable of producing sounds of different physio-physical types, it may use all of them as long as there is but one consonant or one vowel "phoneme" in its system of expression. It should be remembered then that it is quite arbitrary to call the sequence *pa* two *phonemes*. It is true that on pages 239–240 we do not find the word "phoneme," but we see "constituents of this utterance" (p. 239), "optimal consonant," "optimal vowel" (p. 240), all indicating that *pa* is viewed as a phoneme sequence. At this stage, there seems to be *any* room for allophonic variation, unless we are to believe that the child leaves babbling completely behind before it acquires the "nuclear syllable." As long as the allophonic variation exists, phonemic notation must be a matter of choice. It may be tempting to

suppose an "optimal consonant" and an "optimal vowel." If these are chosen or preferred by the young child, there will be reasons for the choice, and the reasons may well be those which the authors put forth. But one is not convinced that this "phoneme" *p* with only one opposite term (the vowel) has no important consonantal allophones. Compare IRWIN, p. 416. After p. 222, there is no question of allophones, though the distinction of allophonic *vs.* phonemic was one of the great achievements in *Kindersprache* . . . (cf. W. F. Leopold in *For Roman Jakobson*, p. 286).

The linguists of this century, have often kept diachronic and synchronic phenomena rigorously apart. We witness, however, that the man who in 1949 wrote: ". . . les changements phonologiques procèdent par bonds" (see N. S. Troubetzkoy: *Principes de Phonologie*, Paris, 1949, p. 318), in the *Manual* asserts that "any sound change, at its proceeding, is a synchronic fact. Both the start and the finish of a change coexist for a certain length of time" (p. 249). This is made possible by "code switching" either between generations or in "one and the same generation between two stylistic levels."—"Thus synchronic analysis must encompass linguistic changes . . ." This would illustrate what U. Weinreich has termed "the dynamism of synchrony," (*Word IX* [1953], 312).

Chapter 15. *Phonetics and Linguistic Evolution*, by A. MARTINET, pp. 252–273. The author complains of lack of "phonetic likelihood" in what he terms "metachronic equations": Lat *ū* = Fr. *ü*. He warns against the "defeatism" of traditional linguists when they consider shifts without analysing the causality (pp. 252–255). He states that "glossematics has deprived itself of any grasp on evolutionary data," and that it "seems unable to perceive anything but constancies in the successive stages of the same language" (p. 255), which is hardly apt to, and cannot even have been intended to, discourage Hjelmslev.

Instead of the rules of the "interaction between neighbors in the chain," Martinet, in tackling problems in historical phonology, stresses the importance of the "interaction among neighbors in the system" (p. 256). There is "a field where the assumed arbitrariness and randomness of phonetic changes must find limitations" (p. 257).

The meaning of the term "distinctive power" (thrice on pp. 258–259) is not clear to the reviewer. "Functional yield" is explained later (p. 266); we hesitate to take these two terms as synonyms.

In sound change there is more than "a neat conflict between the forces of order and conservation . . . and least effort" (p. 257). And since "the phonological system is a coherent whole" (p. 257), the regaining of balance after having been affected on one point is a far-reaching

procedure which may take centuries, and partly coincide with new effects of unbalance at other points, so that "no system is ever found which is not in a process of reorganization at some point or other" (p. 258).

The use of the term *discrete* "from the mathematical jargon" (p. 259)—"discrete" units are "units whose linguistic value lies in their presence or their absence and is not susceptible to degrees"—at first seems a plus to terminology since the term is opposed to *prosodic, suprasegmental*" (p. 260). However, some of the suprasegmental features (the tones) "have a distinctive function identical in its principle with that of phonemes" (p. 260), which are discrete, so that all it amounts to is saying that some suprasegmental features are phonemic.

"Accent, another prosodic unit," is declared to be occasionally discrete: Spanish *páso*, 'I pass' ~ *pasó*, 'he passed' (p. 260), "but this discrete character of accent differs from that of phonemes in so far as accent will never be absent from a complete utterance in an accent language . . ." (p. 260). Within the frame of an utterance with more than one accented syllable, accent will be "susceptible to degrees," and thus it is no longer discrete, as, for instance, *b* in *back*, which is always discrete no matter with how much energy it is pronounced. To solve this difficulty in describing the "discrete character of accent" (stress, p. 260), we would propose to consider accent in itself as *not discrete*, and derive a discrete character from *accent patterns* only, i.e. from the whole comportment of stress features in the syllable sequence: *páso* ~ *pasó*, or English *éxport* ~ *expórt*, and by saying that if accent can be distinctive, it is not because it is either absent or present ("discrete"), but because it has certain distributional properties within the word boundaries; it comes before or after an unstressed vowel. These *kinds of distribution* ('stress patterns') are discrete; they are present or absent, and not susceptible to degrees. That would restore the consistency of the statement about the linguistic value of discrete units, and it would mean that we treat strong and weak stress as we do high and low pitch, giving them linguistic importance according to their relationship in patterns ("tones").

Intonation is declared to exert "a decisive influence" upon tones, thus affecting "some fully arbitrary units of the phonological systems" (p. 261). It is also said that "an influence of intonation on phonemes is possible, probable in certain cases, but difficult to observe" (p. 261). Any influence on tones or on phonemes *as such* would have to be considered on the phonological plane. It is difficult to see, therefore, how, speaking of tones, the result of an influence by intonation can be more, or less, than neutralization. As to the influence from intonation on segmental phonemes, it seems

probable that only their realization (substance) can be meant, not their function.

Of considerable interest is the section dealing with phonemic mergers, and compensation for merging (pp. 262–267). Another section discusses the economy of combining articulatory features to make phonemes; this system is somewhat different from that of the *Preliminaries* and Chapter 14.

In a span of fourteen lines (pp. 268–269), the word "articulation" occurs nine times. The first five times it means something like "splitting," e.g. "articulation of phonemes into distinctive features" (p. 269), compare Barney and Dunn, pp. 193–195; the last four times it must be taken to mean something not far from "action" (of the speech organs), compare Chapter 10.

b. Special Part

The General Introduction to this Special Part starts with the statement that "languages may show a certain degree of relation at various levels" (p. 277). This probably is to say the least of it. The authors of the chapters on Romance, Germanic, Slavonic, Finno-Ugrian and Semitic languages have accepted certain limitations and consented to adhere to a general scheme of presentation. The justification of any such model is not easy to see. Three sections are to deal with the following matters: I. "The specific articulation and articulation basis," II. "The connection between the various speech sounds and . . . the formation of the syllable," III. "Rhythmic and melodic patterns."

On p. x, the editor regrets not to have been able to offer a place to other languages. It seems to the reviewer that instead of trying to present abstracts of pronunciation rules for a long series of languages which must necessarily be only a selection, it would have been wiser to take one single language and give a fairly extensive presentation of its pronunciation for the purpose of showing how phonetics is applied in the description of a language. This language could be the one in which the book is written. One of the difficulties which would then be overcome is the divergent use of such terms as "diphthong," pp. 284, 294, 305, 319, or the way in which chapters 16–19 deal with the "basis of articulation." The Romance languages, in this respect, are compared among themselves, the Germanic languages are mostly considered in relation to other language groups; so are the Slavonic languages, the articulation basis of which "binds them as a whole" (p. 302). As characteristic features of the articulation basis in Slavonic languages are mentioned: 1. expiratory breath-stream, 2. "a tendency to avoid certain difficulties of articulation (except palatalisation)",

3. nearly always lax vowels, which are neither open nor close, 4. several phonetic features of the consonant system: great number of palatalized consonants; consonant clusters not "particularly frequent," voiced and voiceless consonants kept distinct, lack of aspirated and interdental consonants, 5. generally medium force stress (except in Russian), generally descendent intonation curve. In the description of Finnish and Hungarian, "articulation basis" comes nearest to what Sievers called "Operationsbasis" (*3Grundzüge*, p. 103–104); in Sovijärvi's expression, the "preparatory position" (p. 314). Sovijärvi gives ethnological facts to account for differences in this "preparatory position." The statement of Olaf Broch is confirmed: what is called "articulation basis" is "ein etwas nebelhafter Begriff" (*Slavische Phonetik*, 1911, p. 109), and therefore not very useful.

Chapter 16. *The Romance Languages*, by G. HAMMARSTRÖM, pp. 279–288. This sketch could be the notes to a fine lecture on Romance Phonetics, where details can only be hinted at. Important features stand out well; in some places, one regrets that space has not permitted the author to discuss his views. This applies to Fr. *ouaille*, which "should be interpreted as u+stressed $a+i$" (p. 284). See also line 6 from bottom, and p. 286, middle. The characterizations "(extremely) slight," "almost the same" are not of much use in phonetic description, cf. p. 285, and also pp. 305 (Hála), and p. 313 (Sovijärvi). The curious reader will miss a description of the difference mentioned.

Chapter 17. *The Germanic Languages*, by L. F. BROSNAHAN (Introduction, Segmental sounds. Combinations of segmental sounds), and O. VON ESSEN (Rhythm and melody in Germanic languages), pp. 289–301.

Page 289: "Generally speaking the tension of the articulatory muscles is moderate." This fact, if it deserves the name, is taken to account for diphthongization in English and Dutch, "although in Scandinavian languages monophthongs have developed from former rising and falling diphthongs." Little information can be gained from such statements. Page 290: Dutch is said to have two voiced "dentolabial" fricatives (not to the exclusion of *f*, though the wording would suggest it). The difference of the symbol values *v* and ʋ is not made clear, either in the text or on pp. xiii–xiv. Page 291: "The uvular trilled ʀ is common in German, Dutch and Danish, though in the last a uvular fricative ʁ, is commonly used." As regards Danish, the statement of *uvular trilled* ʀ is surprising, and if it is "common," how is one to understand that a *uvular fricative* is "commonly used"? Are they both common, and if so, are they equally common? Page 291: The term "distinctive phoneme" (for laryngeal stop in Danish) is a good example of the manifold usage of "phoneme." It could be ex-

pected that authors using it in a *Manual* would not deviate from the general idea of distinctiveness in this unit of expression, so that "phoneme" would not need the epithet of "distinctive." The author has found (p. 292) that "in German fortis articulations such as *t*, *k*, *f* are distinctly voiceless. Now the rather questionable statement that American English commonly has voiced *t*, *k*, *f* throughout their duration, "especially between vowels," and the view that "Dutch and British English take up intermediate positions in this respect", lead the author to assume "an East to West decrease in the degree of voicelessness of fortis consonants." We would not subscribe to a theory including the North Atlantic in a language area. But it could also be objected on other grounds that speakers of American English seem to voice their stops only intervocalically if one does not include cases of voicing before voiced consonants, which need not concern us here. Intervocalically, voicing may not be too infrequent in "normal relaxed" German, e.g. "ich hatte" pronounced with voiced *t*. The reviewer is prepared to submit one case of such voicing in German, recorded on sound tape. On page 292 the author mentions *f* besides *s* as a consonant which checks the aspiration of following fortis stops in all the Germanic languages except Dutch. This is a good point. But it cannot be seen what is meant by "Elsewhere the degree of aspiration varies, probably in close correlation with the degree of fortisness in the articulation" (p. 292). Page 293: The table of "vowels and consonants present in Germanic languages" shows 15 Norwegian vowels, among them *ə*, which is "present" also for English, Dutch, German, Danish, and Swedish. The table can therefore hardly be considered to give *phonemes* like the one for Romance languages (p. 280). If there are more "vowels" than there are vowel phonemes in the table on p. 293, the "consonants" listed for Norwegian are too few to represent all the consonant phonemes in the most frequently described (Eastern) variety of that language; the retroflex (or supra-alveolar) series is lacking. The same would apply to Swedish. The Scandinavian languages, according to the table, have no labiodental voiced fricative: [v], but one which is bilabial: [w]. The symbol *q* for some Danish sound is unexplained, and so is one of the *r*-sound symbols, p. 294, middle. For reasons stated, p. 289, note, no reference is made to literature. Such references would have been especially welcome on account of the many unorthodox views of the author.

The third section, Rhythm and melody in Germanic languages, seems to reveal that stress patterns are more typical than pitch patterns. Page 299: The term "supplement-question" ("Ergänzungsfrage") does not seem to be a happy choice, since the *question* is no *supplement* in itself; nor does it require any supplementary question. The idea of "supplement" is that

the question calls for an answer which supplements the ideas involved in the question itself. The symbols of stress and pitch comply with the requirements of *the sufficient and the necessary*. On a few points corrections could be made; p. 298, bottom, "*U(ge), da(ge)ne hedder*," p. 299, "how many miles is *it*" (stress mark). This fine outline of the question of intonation reveals many common features in the languages treated.

Scandinavian readers may miss a brief sketch of toneme realization. It is known, however, that the dialect differences in tonemic pitch patterns has led the author to suspect that in European languages the word tones result from contextual patterning and do not belong to the "internal structure of the language" (*Allgemeine und angewandte Phonetik*, 1953, p. 124).

Chapter 18. *The Slavonic Languages*, by B. HÁLA, pp. 302–311. Page 303: The phonetic transcriptions of those affricates "whose second elements are sibillant [sic]"—of what kinds are the others?—and of "*ń*" and *ř* (p. 304), do not conform to the symbols given at the beginning of the book (p. xiv). The reader must wait until p. 306 to understand what sounds these symbols represent. It would be difficult to tell what is cause and what is effect in the case of missing labialization of front vowels in the Slavonic languages (p. 303). The author says that labialization is "caused by the fact that the activity of the lips is small as compared to German and French, . . ." The same cause-and-effect explanation puzzles the reader in the question of palatalization: "The main reason seems to be that the speakers of the languages concerned are in the habit of pressing the mass of the tongue into the palate" (p. 304f.). It is not easy to understand what is meant by the articulation of the tongue being "always dorsal": this would seem to be characteristic of the "articulation basis," but is not mentioned in that connection. One needs to know what the author considers as the opposite to "dorsal"; is it "retroflex"? Clearly, it is not "blade" (coronal). The term "loose connection" (p. 307) leaves the reader in some uncertainty, since it seems to be used also for the transition consonant-vowel, while this and similar terms ("open contact," p. 229) is with very few exceptions (Forchhammer) used to specify the vowel-consonant transition only. No examples of "loose" or "closer connection" are given, here or elsewhere in the *Manual*. On page 307 the reader will probably not be able to see how Czech *sekra* furnishes an example of *glide* in comparison with *sekera*, or how Russian *Pavlovič* (according to the author not on account of "the quick speed of speech" but because of "a strong stress") gives a form with a glide: *Palč*; where is the glide? Page 307: Some words fail to add to the clarity of the phonetic description. One of them is "special," as in the sentence, "Phonetic changes of con-

sonant combinations do not offer anything special." If it means that something like the changes in Slavonic languages are found elsewhere, this can hardly be a reason for not mentioning them. Compare the following statement: "As far as emphasis and dynamical arrangement of stress-groups into higher logical units (sense-groups) is concerned, the Slavonic languages have no special acoustically expressive means" (p. 310). Another such confusing word is "difficult," used about consonant clusters (p. 307), or the expression "certain difficulties of articulation," already mentioned (p. 302). Readers who do not speak Slavonic languages cannot be expected to know what to native speakers is difficult to pronounce. The cases mentioned of "intrusive sound" (p. 308) can hardly have much interest except in language history; they do little to illustrate the phonetic system of Slavonic languages, either as a whole or as compared to each other. The same thing must be said of the passages on Dilatation and Dissimilation in loan words of long standing (p. 309). The statement: "Reciprocal influence of sounds in connected speech is manifested in the same manner as in the non-Slavonic languages," seems rather broad, considering the title of the section (Articulation in the flow of speech) in which it appears. Here a number of phenomena of this type are treated separately. Voice assimilation, for instance, would not seem to play the same part "anywhere else" (p. 308). A good formulation: assimilation of voice (here: regressive voicing) "only exceptionally takes place before those consonants that have no voiceless opposites" (Bulg. *pi∫me—piʒme*). The last section, Intensity and pitch in Slavonic languages, contains good information on the structure of rhythm groups, but only a mere hint is made at phonemic pitch patterns in Serbocroatian and Slovenian (p. 310). The formulation that in Russian "there is a great difference between the written and the spoken word" will scarcely satisfy many readers. Finally, one wishes that translations had been given of the examples from the Slavonic languages.

Chapter 19. *The Finno–Ugrian Languages*, by A. SOVIJÄRVI, pp. 312–324. Some remarks on this chapter have already been made in connection with other contributions to this part of the *Manual*. Page 320: The case of *nyt on* pronounced *nü-ton* or *nüt-ton* is interesting for its parallel in French (*il a* becoming *il la*, or, under certain circumstances, *i la*); compare also p. 283: Italian *da me* 'from me' sounding like *dam͡me*. Page 322: "Finnish is a centralized language . . ." This statement, which in France or in Norway might be taken to express that there is a very influential norm, needs an explanation, or a reference to Alfred Schmitt, who coined the term, and kept it (1924, 1953). It is of great interest to note (p. 323) that Finno–Ugrian shares with Germanic and Romance languages a non-typical pitch pattern in interrogative sentences beginning with an

interrogative pronoun; in the literature occasions are rather rare where resemblances between unrelated languages are brought out.

Chapter 20. *The Semitic Phonetic System*, by W. LESLAU, pp. 325–329. The author has chosen to give in tabular form the Proto-Semitic sound system only. We get details of sound development in Semitic languages, but no general survey of modern "phonetic systems." The author conscientiously sticks to the task of describing the "basis" and the development of sound systems, even with phonemic evaluations, but the whole means little more than tracing the history of single phoneme realizations. The questions of sandhi effects, of syllable formation, and of "rhythmic and melodic patterns" (pp. 277–278) are not discussed. Still, viewed as an historical sketch, this chapter appears to have great merits for fullness of detail and objectivity of presentation (p. 326). Two misprints: It is highly confusing at first that the symbol for the "emphatic" interdental is replaced by that of the corresponding dental (p. 325). Page 326, line 4 from bottom, should probably read: "of emphatic: lenis: fortis seems to" etc.

Chapter 21. *Notes on African Tone Languages*, by A. BURSSENS, pp. 330–335. This survey of toneme realization in African languages is probably meant by the editor as in a way representative of writings on tone languages in general. The notes are numerous, but rather short. They do, however, do justice to chief aspects of tonology, such as level *vs.* gliding pitch, grouping of tones in tonemes, and "lexical" *vs.* "grammatical" tonemicity.

Part E. Phonetics in Its Relation to Other Sciences

Chapter 22. *Phonetics and Pathology*, by R. LUCHSINGER, pp. 339–363. This chapter deals with the "particular scientific aspect of phonetics" which the author calls "experimental methods of recording cases of pathological speech and voice sounds" (p. 339). Section 1 is concerned with the "monotony of speaking voice" (p. 339). Experiments are reported which, in the opinion of the author, show a "striking difference between melody curves of epileptics and of normal speakers" (pp. 340–341). However, the "rich fluctuations from one wave to another" (p. 341) and the findings of Schär and Zwirner (mentioned by the author) that the monotony heard may be just imaginary, seem to indicate that the example is not very illustrative of recording methods. Of more interest are the reports of acoustic experiments in cases of dyslalia (Section 2), and laryngectomy (Section 3).

Section 4, Recording the respiratory output and the intensity in phonation, deals with pneumotachography (respiration timing) and sound pressure gauge for self-observation of the sound pressure in the singing voice.

In Section 5, The analysis of sibilants and sigmatisms by means of electro-acoustic devices, the author examines the effects of hearing loss on pronunciation. It seems possible from the description of the instruments to place in the hands of subjects with defective sibilants electrographic devices for facilitating speech rehabilitation (so called "*s*-indicators").

The last section of this chapter deals with Electro-encephalographic studies in clutterers and clutterer-stutterers. It is of interest to note that pure stuttering, "a psychoreactive, conflict-conditioned, chiefly tonic disorder, mostly appearing in later life—is nearly always attended, as far as we could see, by a normal E.E.G." (p. 362).

Chapter 23. *Phonetics and Sociology*, by A. SOMMERFELT, pp. 364–371. The author discusses social aspects of "distinctive traits" (not the same as the "distinctive features" of Jakobson-Halle). Different alleged causes of linguistic change and reaction to change are mentioned and evaluated. The author finds that the role played by phonemic/phonetic facts in social life is clear. "But the question whether there exists, on the synchronic level, a more precise correlation between phonemic structure and the structure of the social group which uses it, is unsolved and even has to wait for a method of approach" (p. 371). Until such a method is found, one may speculate whether a correlation is possible without causality, which again involves diachrony.

Chapter 24. *Phonetics in Its Relation to Musicology*, by J. SMITS VAN WAESBERGHE, pp. 372–384. This chapter is concerned with the difference between speaking and singing, and with an aspect common to both: the recitative. The first two sections (a, b) will raise doubts on salient points. It is difficult to see how the author can compare "slight nuances of pitch" "when a singer or a choir . . . goes sharp or flat with regard to the key-note on which the song was started" (p. 373), with "the slowly rising or falling pitch of the voice of an orator or a reciter" (p. 374), and how he can "establish this contrast between speech and music: slight differences of pitch are spontaneously perceived and interpreted by the hearer, while in music they hardly impress the ear at all, or, if they do, are not appreciated as differences of pitch" (p. 374). We indeed agree with the author where, surprised at his own statement, "however well accounted for" (p. 374), he reflects that he would be more inclined to expect a nicer perception of pitch in music. *How* well the author accounts for the appreciation of slight differences of pitch in speech as opposed to music, should be mentioned. The half-page argument can be summed up in the following wording (p. 374): "What is of primary importance in speaking is the deviation from the normal, in music it is the normal and not the deviation that demands

attention." Phonemicists, even phoneticians, would refuse to support this view; all recognition, whether in speech or in song, requires the observation of norms.—As to rhythm in speech *vs.* time in music, "we come to the same result as was arrived at above for pitch: man has a very delicately balanced power of perceiving the first but reacts only indifferently as regards time in music" (p. 376). The reason is again considered to be "a quite different perceptive attitude" (p. 377). This becomes understandable if one takes into account that, according to the author, the rhythm of speech "embraces various elements: length of sounds, pitch, stress and pronunciation of consonants and vowels" (p. 376). The phonetician, with some confusion, here recognizes the whole of phonetic substance.

Under the heading *c. From speaking to singing*, the result of the following experiment is discussed: A reciter, non-musician, changes gradually from "ordinary recitation" to music (song), and a musician (singer) slowly changes from singing to speech: "the two performances will not meet at any point" (p. 380). However, no conclusions are drawn. Instead, we get an outline of the recitative, interesting enough in itself, and, perhaps, of importance to phonetics as a kind of parallel to speech melody.

Chapter 25. *Phonetics in Relation to Aesthetics*, by A. W. DE GROOT, pp. 385–400. This contribution is valuable as an analysis of methods in the investigation of literary art (verse), but does not go far into the question of the aesthetic qualities of speech. It cannot be seen, therefore, that the chapter throws much light on phonetics.

Part F. Phonetics and the Origin of Speech

This part is devoted to the development of speech, first in the child (Chapter 26) and then in mankind (Chapter 27). Finally, a chapter called General Semantics compares linguistic and "sublinguistic" expression.

Chapter 26. *Phonetical Description of Speech Development in Childhood*, by O. C. IRWIN, pp. 403–425, is first of all remarkable for the fact that mention is nowhere made of Jakobson, at least not directly. The author raises the question: "Are there any laws in the early premeaning stages of speech development?" It is stated that "there are laws of development during the period of infancy" (p. 424). The next question discussed is, "Can the laws be used practically in the clinic to diagnose retardation?" It is found that the laws are different in the speech of handicapped children, which permits an early diagnosis of speech handicap. "Concerning the remedies for retardation in infant speech it should be emphasized that further research on the . . . two problems is needed for their formulation" (p. 404).

Readers will soon become aware that the author uses the term "phoneme" in the sense of "speech sound," cf. text fig. 1 (p. 412): "Development of phoneme types" and p. 418: "speech sound development." The author arrives at the conclusion that the labials *p*, *b*, and *m* are not "the consonants which are first mastered" (pp. 410, 416). The meaning of "mastered" in this connection clearly has nothing to do with phonemic function, as it would according to *Kindersprache* . . . Understood rightly, however, it is interesting to note the finding reported on p. 412: "While the average young child increases its mastery of the phonetic elements with age, it does so at a decreasing rate." The investigation of the development of "phonemes" in the baby enables the researcher to determine its "phonetic status," which seems a very appropriate term. Unfortunately, it is probably used indiscriminately together with "phoneme status" (pp. 414, 415). For specification of speech sounds "mastered," vowel and consonant *profiles* are used. Phoneticians will note that *h* is ranked as glottal, along with the laryngeal stop. This is in itself hardly worthy of mention, since in consonant charts the symbol *h* seldom leaves the space for "laryngeal fricative," where it seems to be locked up by tradition. What gives importance to the point here, is the fact that the author uses the expression "glottal endowment" to describe the share of glottal sounds produced by normal and cerebral palsied children, and applies it as a "further demonstration of the terrific retardation in this form of brain damage." But is *h* glottalized?

Chapter 27. *Development of Speech in Mankind*, by E. Buyssens, pp. 426–436, differs radically from the foregoing chapter by confining itself to probabilities. Whether this fact makes a reviewer less critical, or whether the probabilities seem as well founded as one can wish, the fact remains that this is fascinating reading. Though the author states that "we must study the sign before the phoneme" (p. 427), this does not invalidate the attempts to follow the development of *parole* in childhood (Chapter 26). The author finds that "the pun is the clue to the origin of the articulate language" (p. 429). By "dividing the indivisible," "the interjections came to be regarded as being composed of functional parts, i.e. of signs: interjections became articulate sentences" (p. 430). (Even "mew!" can subsequently be analysed phonemically, cf. Hammarström p. 284.) By the same act of dividing, "man learned to dispense with intrinsic significance . . ." (p. 432); ". . . he could also devise new signs . . ." (p. 432). But first pronunciation had become articulate: phonemic systems were created. At this point (p. 433), the author has the courage to call his own definition of the phoneme, "the definition of the phoneme": "The phoneme is the smallest unit that can be both common to two signs otherwise different, and different in two signs otherwise identical."

(Published for the first time in *Cahiers Ferdinand de Saussure* VIII [1949], 54). A small number, and a system, of such units is made necessary by the large number of signs, since *more signs* will mean more signs relying for their interpretation on readily identifiable constituents. The more numerous the phonemes, the less clear the differences. "In general, we can say that in every system of phonemes there is room for more phonemes than actually used" (p. 435).

This fine contribution to the *Manual* ends with some remarks on attempts at "articulating" (one might perhaps here also call it "sign-splitting") intonation (Chinese and English) and stress (French) (p. 436). As to the idea of sign-splitting in Chinese tones, it should be noted that this is a parallel to the view of von Essen, mentioned under Germanic Languages, above. In both cases, the point seems to lack support in language history; "tones" are basically different from intonation.

Chapter 28. *General Semantics* (in the special sense of "non-phonemicizable, unarticulated expression"), by F. TROJAN, is the shortest of all the chapters (2½ pages). Yet it reminds the phonetician that he has more than phonemic realizations in his field of research.

This *Manual*, then, ends as it began, by maintaining the autonomy of Phonetics in relation to Linguistics.

Non-phoneticians may be better judges of the success of this enterprise, which is not made for the benefit of specialists. Phoneticians will read most of the chapters with interest and find useful knowledge especially in the parts on Basal Sciences and Phonetic Research.

University of Bergen
Norway

NOTICE

The founding of *The Journal of Auditory Research*, an interdisciplinary non-profit quarterly devoted to the scientific study of hearing, has been announced. The new journal will cover the fields of psychoacoustics, otology, audiology, neurophysiology of audition, speech and communications, auditory aspects of human engineering, musicology, instrumentation for hearing research, and all other aspects of audition. Dr. J. Donald Harris, of the Medical Research Laboratory, New London, Conn., will be Editor, and Dr. James Jerger, of Northwestern University, will be Associate Editor. The Editorial Policy Board consists of Drs. Norton Canfield, Raymond Carhart, Stacy Guild, Henry L. Haines, Fred Kranz, Al Liberman, and E. Glen Wever. The new journal will be published by the C. W. Shilling Auditory Research Center, Inc. of Groton, Conn. Initial publishing costs have been defrayed by a grant from the Beltons Institute for Hearing Research.

Manuscripts and preliminary reports on all aspects of hearing are invited. They should be addressed to the Editor, 348 Long Hill Road, Groton, Conn. Advance subscription will be $3.00, regular subscription, $5.00.

L. C. N. Y. Publications

MONOGRAPH SERIES

The *Linguistic Circle of New York* publishes occasional Monographs as supplements to various volumes of WORD. These Monographs are distributed gratis to members *at the time of publication.* They are available thereafter only on specific order and at the prices listed.

Monograph No. 1, Supplement to Volume 7, WORD
KARL HEINRICH MENGES—The Oriental Elements in the Vocabulary of the Oldest Russian Epos, The Igor Tale ..... *out of print*

Monograph No. 2, Supplement to Volume 9, WORD
YURY SERECH—Problems in the Formation of Belorussian **$2.50**

Monograph No. 3, Supplement to Volume 12, WORD
ANNA GRANVILLE HATCHER—Theme and Underlying Question; Two Studies of Spanish Word Order **$2.50**

SPECIAL PUBLICATIONS

1. URIEL WEINREICH—*Languages in Contact* (1953), xiv+146 pp. ..... *out of print*
2. *Linguistics Today* (1954), 280 pp. ..... *out of print*
   Equivalent to WORD, vol. 10, no. 2–3, but with hard-cover binding and separate pagination.
3. *The Field of Yiddish; Studies in Language, Folklore, and Literature*, ed. URIEL WEINREICH (1954), ix+317 pp. **$5.00**
   For members and subscribers (Please mark your order "Member Rate") **$4.00**

Complete sets of back numbers of WORD are no longer available. A list and quotation of those still in print will be sent on request. Information about reprinting will be found on the outside of the back cover.

Write to: JOHN P. HUGHES, St. Peter's College, 2641 Hudson Blvd., Jersey City 6, N.J.

ANNOUNCING

A SUBSCRIPTION CAMPAIGN

FOR REPRINTING BACK NUMBERS

The Linguistic Circle of New York is happy to announce that it has completed arrangements for reprinting the Monographs, Special Publications, and back numbers of WORD that are presently out of print, and virtually unobtainable.

Reprinting will be on a subscription basis. Those desiring back numbers or one of the Publications should send their orders and remittances to the Treasurer of The Linguistic Circle of New York (John P. Hughes, St. Peter's College, Jersey City 6, New Jersey). All orders will be acknowledged, and all remittances held in a special fund. When a sufficient sum is in hand for reprinting any number, it will be sent to press. If by December 31, 1962, sufficient orders have not been received to warrant reprinting any particular number, all remittances for it will be refunded as soon as possible after that date.

Following is a list of the publications presently out of print. Where not otherwise indicated, each single number is priced at $3.00, each double number $6.00, i.e., $9.00 for the ordinary volume of three numbers. (This includes postage anywhere in the world.) In cases where a bill is required, a pro-forma invoice will be forwarded on request.

WORD

Volume		*Number*	*Volume*		*Number*
II	(1946)	1–3	VIII	(1952) (No. 4=*Slavic Word* #1)	1–4
III	(1947)	1–3			
IV	(1948)	1–3	IX	(1953)	2
V	(1949)	1–3	X	(1954) (No. 4=*Slavic Word* #3)	1–4
VI	(1950)	1–3	XII	(1956)	1–3
VII	(1951)	1–3	XIII	(1957)	1–2
VII	*Supplement* (*Monograph* #1)		(All subsequent numbers in print).		

PUBLICATIONS

Languages in Contact, By Uriel Weinreich $5.00

Linguistics Today (A separately paged and bound reprint of WORD X, No. 2–3) $5.00

MADE AND PRINTED IN GREAT BRITAIN BY
WILLIAM CLOWES AND SONS, LIMITED, LONDON AND BECCLES

VOLUM

JOU

ARTICL
Tranche
"Modal
Remark
Stress-D
Some Pr
An Intr
DISCU
Linguis
REVIE
Kennet
Part
Eva Siv

Mauric
Elliott
Holger
Linguis
Heinz-
PUBL
LING

T